McMaster University
Inspiring Innovation and Discovery

Guideline development using GRADE

September 9, 2011

Holger Schünemann, MD, PhD

Professor and Chair, Dept. of Clinical Epidemiology & Biostatistics
Professor of Medicine
Michael Gent Chair in Healthcare Research
McMaster University, Hamilton, Canada

Faruque Ahmed, MD, PhD

ACIP
CDC

Rebecca Morgan

CDC

CANADA
Sudbury
Sault Ste. Marie
North Bay
Montréal
Ottawa
Eau Claire
Wausau
Cheboygan
Petoskey
Alpena
ONTARIO
Cornwall
Dover-Foxcroft
Bangor
WISCONSIN
Green Bay
Owen Sound
Brockville
Burlington
M. E.
Manistee
Kingston
Sheboygan
Lake Huron
Watertown
Lewiston
MICHIGAN
Bay City
N.H.
Toronto
Madison
Milwaukee
Flint
Utica
Manchester
Lansing
Hamilton
Syracuse
Boston
Kalamazoo
Buffalo
Rockford
Detroit
Erie
Providence
Chicago
Toledo
Cleveland
New Bedford
Peoria
Fort Wayne
Lima
Normal
Kokomo
Akron
Muncie
OHIO
Indianapolis
Columbus
Charleston
Dayton
St. Louis
INDIANA
Athens
Cincinnati
W. VA.
Washington, D.C.
ILLINOIS
Louisville
Farmington
Charleston
Charlottesville
Evansville
Lexington
Richmond
Paducah
MO.
Roanoke
Somerset
KENTUCKY
VIRGINIA
Atlantic
Ocean
Clarksville
Nashville
Kingsport
Suffolk
Virginia Beach
Jackson
Winston-Salem
High Point
Raleigh
Knoxville
Lawrenceburg
TENN.
Charlotte
Florence
Greenville
Huntsville
Spartanburg
Tupelo
NORTH CAROLINA
New Bern
MISS.
GEORGIA
SOUTH CAROLINA
ALABAMA
Birmingham
Atlanta
Myrtle Beach
Sumter

The Department of Clinical Epidemiology & Biostatistics at McMaster

History

- 1967 – Founded by David Sackett
- 6 chairs since
- Instrumental in specialty of Clinical Epidemiology, origin of “Evidence-Based Medicine”

People

45 full time and joint faculty

~ 120 associate & part time faculty; 19 emeritus

~ 180 staff

~ 200 PhD and Master students

Agenda

McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

Agenda

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

What is a guideline?

- "Guidelines are recommendations intended to assist providers and recipients of health care and other stakeholders to make informed decisions. Recommendations may relate to clinical interventions, public health activities, or government policies."

WHO 2003, 2007

Guideline development Process

Key issues

1. Guidelines for guidelines
2. Priority setting
3. Group composition and consultation process
4. Managing conflicts of interest
5. Group processes
6. Determining which outcomes are important
7. Deciding what evidence to include
8. Synthesis and presentation of evidence
9. Grading evidence and recommendations
10. Integrating values and consumer involvement
11. Incorporating considerations of cost-effectiveness, affordab implications
12. Incorporating considerations of equity
13. Adaptation, applicability and transferability
14. Reporting guidelines
15. Disseminating and implementing guidelines
16. Evaluation

Health Research P

Review

Improving the use of resea introduction

Andrew D Oxman*[1], Atle Fretl

Review

Improving the use of research evidence in guideline development: 1. Guidelines for guidelines

Holger J Schünemann*[1], Atle Fretheim[2] and Andrew D Oxman[2]

Published: 21 November 2006

Health Research Policy and Systems 2006, **4**:13 doi:10.1186/

This article is available from: http://www.health-policy-syste

Open Access

McMaster University
Inspiring Innovation and Discovery

Working with evidence

- For key recommendations:
  - Search for and retrieve all available evidence
  - Identify relevant SRs
  - Formally assess quality of evidence
  - GRADE (systematic and transparent approach)

The scope

- Small is beautiful (S. Hill)
- Who is the target user of the guideline
- Who it applies to
- What is covered?
  - Eg diagnosis and treatment of diabetic retinopathy
- Develop key questions (<20…..)

McMaster University
Inspiring Innovation and Discovery

What healthcare workers want…

- A guideline is not a textbook or a cookbook
- To KNOW that the guideline is evidence based
- But not necessarily all of the evidence…
- To have it easy to use and accessible
- Clear recommendations (more on that later)

Who should develop guidelines?

- One systematic review (Murphy et al. 1998)
- Composition of panel influences recommendations
  - Members of a specialty are more likely to advocate techniques that involve their specialty
- Balanced groups
  - Select the appropriate group leader
- Necessary technical skills
  - including information retrieval, systematic reviewing, health economics, group facilitation, project management, writing and editing
- Include or have access to content experts
- No SR on how to obtain consultation, but logical reasons support this
- Up to 15 members

McMaster University
Inspiring Innovation and Discovery

Group composition

- „Include all who are affected“
  - To identify the right questions
  - To identify areas of suboptimal care
  - To identify feasibility of recommendations
- Consequences
  - Definition of Standards of Care
  - Ownership to improve implementation

Expertise needed in the group

- Medical content:

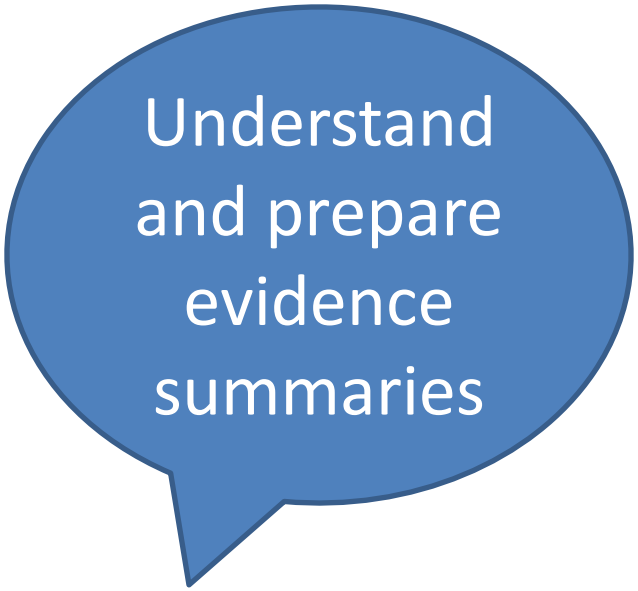
  health care professionals

- Values and preferences:

  patients / carers / community

- Methods and support staff:

  ‚technical' professionals, e.g. epidemiologists, health economists, administrative support



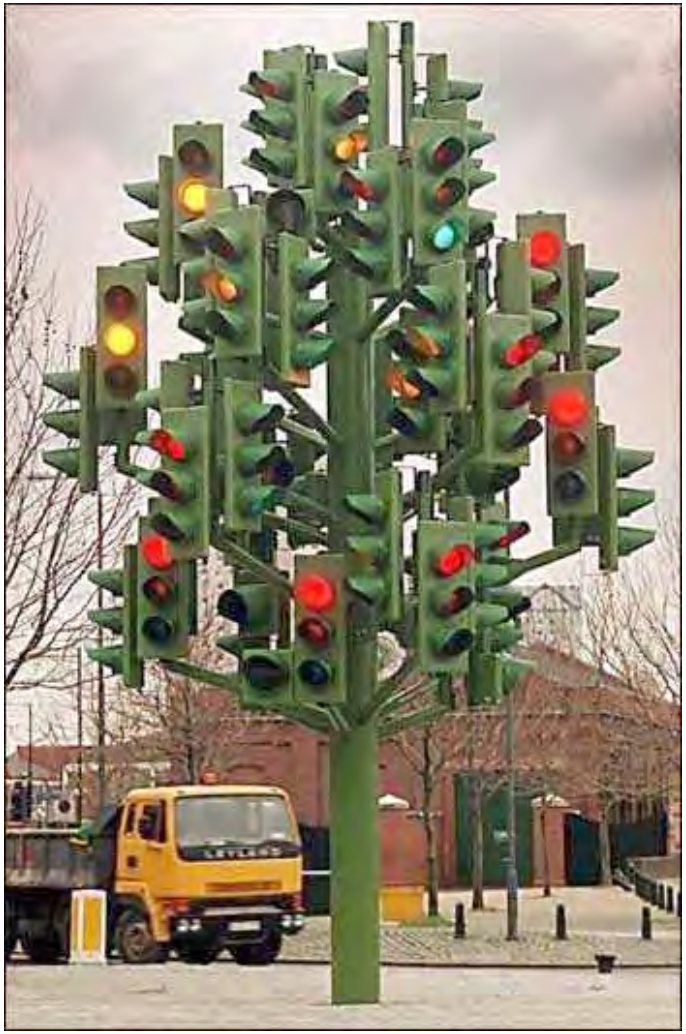
Which approach?

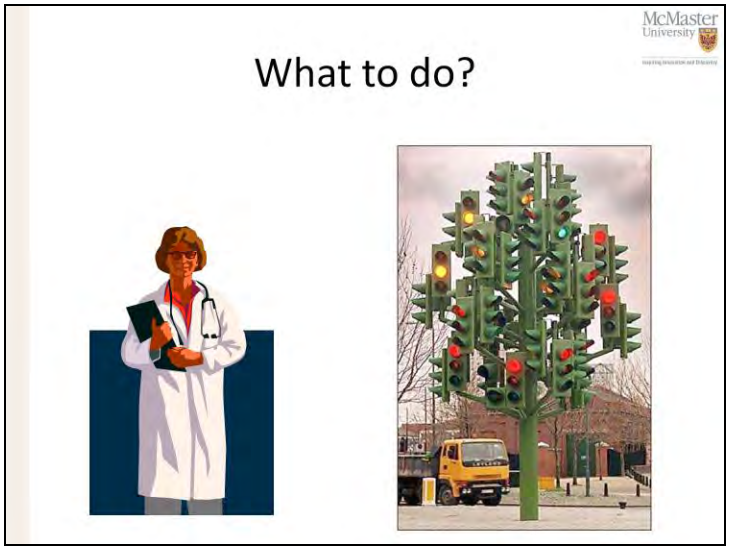
Recommendation for use of oral anticoagulation in patients with atrial fibrillation and rheumatic mitral valve disease

Evidence	Recommendation	Organization
• B	Class I	➢ AHA
• A	1	➢ ACCP
• IV	C	➢ SIGN

McMaster University
Inspiring Innovation and Discovery

What to do?





This information is like standing in front of this particular traffic light, a health care decision maker would not know what to do. The analogy to a traffic light and recommendations is actually a very helpful one as the green light could be indicated or interpreted as implementing a recommendation without much thought, the yellow light depending on where you live in the world would indicate that you should think very carefully and in most other places in the world a red light would indicate that you should stop doing something or you should stop.

GRADE

McMaster University
Inspiring Innovation and Discovery

Working Group

Grades of Recommendation Assessment, Development and Evaluation

- Aim: to develop a common, transparent and sensible system for grading the quality of evidence and the strength of recommendations (over 100 systems)
- International group of guideline developers, methodologists & clinicians from around the world (>200 contributors) – since 2000
- International group: ACCP, AHRQ, Australian NMRC, BMJ Clinical Evidence, CC, CDC, McMaster Uni., NICE, Oxford CEBM, SIGN, UpToDate, USPSTF, WHO

CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005, AJRCCM 2006, Chest 2006, BMJ 2008

McMaster University
Inspiring Innovation and Discovery

GRADE Uptake

- World Health Organization
- CDC-ACIP
- Allergic Rhinitis in Asthma Guidelines (ARIA)
- American Thoracic Society
- American College of Physicians
- European Respiratory Society
- European Society of Thoracic Surgeons
- British Medical Journal
- Infectious Disease Society of America
- American College of Chest Physicians
- UpToDate®
- National Institutes of Health and Clinical Excellence (NICE)
- Scottish Intercollegiate Guideline Network (SIGN)
- Cochrane Collaboration
- Infectious Disease Society of America
- Clinical Evidence
- Agency for Health Care Research and Quality (AHRQ)
- Partner of GIN
- Over 60 (major) organizations

WHO

ATS

IDSA

SIGN
Scottish Intercollegiate Guidelines Network

NHS
National Institute for
Health and Clinical Excellence

bmj.com

THE COCHRANE COLLABORATION®

Formulate question

Select outcomes

Rate importance

Outcomes across studies

Create evidence profile with GRADEpro

Rate quality of evidence for each outcome

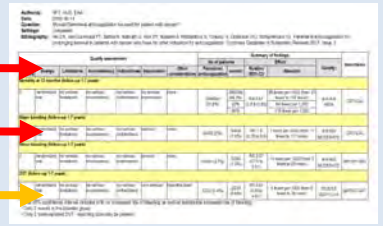
P
I
C
O

Outcome Critical

Outcome Critical

Outcome Important

Outcome Not important



Summary of findings & estimate of effect for each outcome

High
Moderate
Low
Very low

Randomization increases initial quality

Grade down

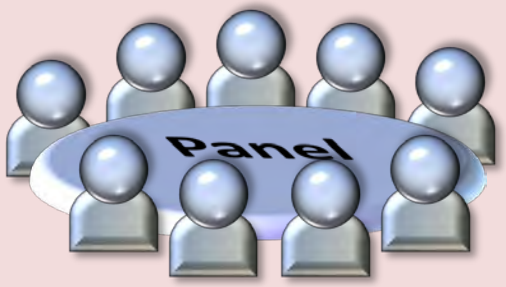
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

Grade up

1. Large effect
2. Dose response
3. Confounders

Systematic review

Guideline development



Grade
overall quality of evidence
across outcomes based on
lowest quality
of ***critical*** outcomes

Formulate recommendations:

- For or against (direction)
- Strong or conditional/weak (strength)

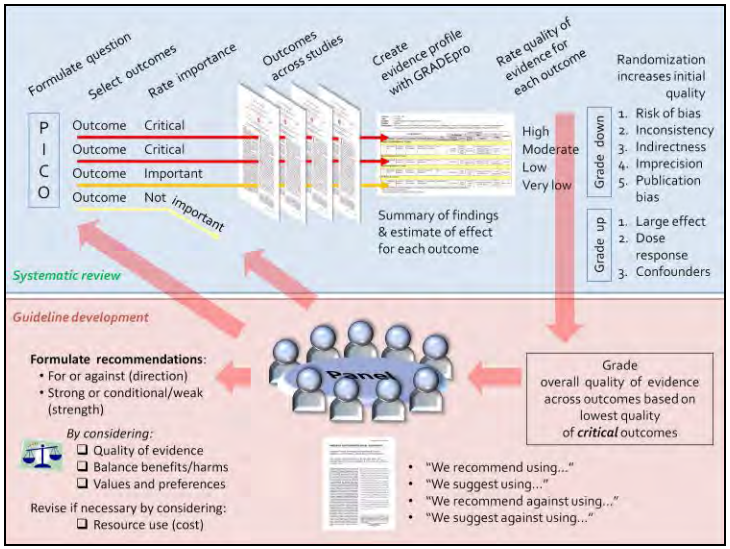
By considering:

- ❑ Quality of evidence
- ❑ Balance benefits/harms
- ❑ Values and preferences

Revise if necessary by considering:

- ❑ Resource use (cost)

- "We recommend using…"
- "We suggest using…"
- "We recommend against using…"
- "We suggest against using…"



This figure demonstrates the ideal process of integrating the GRADE approach into guideline development and the relation between systematic review conduct and guideline development. We will describe this process in an overview first and then describe selected single steps in more detail. It highlights that there is a requirement for a close relation between guideline panels, systematic reviews and those who assess the confidence in the estimates of effect (i.e. the quality of the evidence). It describes that guideline panels should be involved in the development of appropriate healthcare questions according to the PICO framework (reference article 3). The panel is involved in developing these outcomes and selecting the outcomes and in assessing their importance for decision making. This process requires close collaboration of the multidisciplinary panel. Outcomes that are considered critical and important are evaluated in a systematic review. Outcomes that are rated as not important do not have to be considered further. The novelty of the GRADE approach is that the outcomes are evaluated across studies rather than within studies. That is, a different body of evidence may contribute information to different outcomes that are being considered. When an evaluation of the outcomes across studies has taken place evidence profiles using software such as GRADEpro are developed the presentation of this information can either take place in typical evidence profiles or also in the Summary of Findings tables where a detailed assessment of the underlying confidence in an estimate of effect by outcome is then combined with an actual analysis of what the effects are. Those who review the evidence will then grade the confidence in the estimates of effect of a body of

evidence (i.e. the quality of evidence) for each outcome in four categories; high, moderate, low or very low on the basis of 8 factors that either increase or decrease the initial quality. Randomization is considered the best method to protect against bias and confounding and the initial quality of a body of evidence from randomized control trials usually starts as high quality, but there are 5 factors that lower the quality and, usually, for observational studies, 3 factors that increase the quality.

Once all outcomes that are critical for decision making have been evaluated an overall confidence in the estimate of effect to support a recommendation or an overall GRADE of the quality of evidence is assigned. The overall GRADE is based on the outcome with the lowest quality of evidence given that it is a critical outcome. This information is then provided back to the panel.

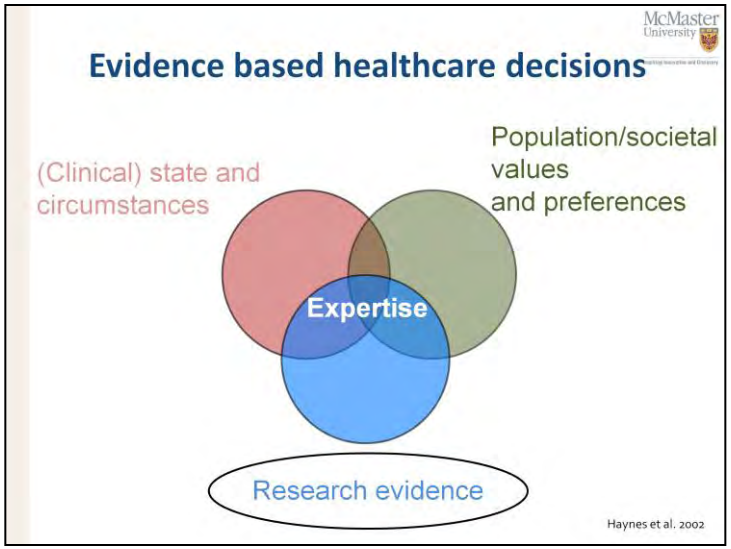
A guideline panel then needs to formulate a recommendation by considering the following 4 factors: the quality of evidence, the balance between benefits and down sides, values and preferences and resource use. A panel will then formulate recommendations in a clear and unambiguous way using standardized wording, such as using the term recommend for strong recommendations and suggest for conditional or weak recommendations or other terminology such as “should” and “may”. Guideline panels will express GRADE’s two directions of the recommendation either for or against an intervention or diagnostic test or strategy and the strength of this recommendation by either determining that it is a strong or a conditional recommendation. Other users of GRADE may use the evidence summarized according to the GRADE approach for health policy decisions.

Evidence based healthcare decisions

(Clinical) state and circumstances

Population/societal values and preferences

Haynes et al. 2002



Fundamentally the GRADE approach is based on the philosophy of evidence based health care decisions that include the integrations of three domains. First it considers the heath state and circumstances, such as where decision making takes place are we dealing with a low income country, a high income country, a primary or a tertiary care hospital, what are the circumstances and the health state that the patient presents with. With the second domain the patient's populations or societal values and preferences how important are certain outcomes for decision making. And the third domain, the actual underlying research evidence. These three domains must be integrated by the use of an individuals or a panels expertise that is required to interpret these three domains and integrate their contribution to health care decision making. When we speak about research evidence it becomes clear that when we integrate research evidence with these other factors that we are implicitly looking for the best evidence.

Confidence in evidence

McMaster University
Inspiring Innovation and Discovery

- There always is evidence
  - "When there is a question there is evidence"
- Better research $\Rightarrow$ greater confidence in the evidence and decisions

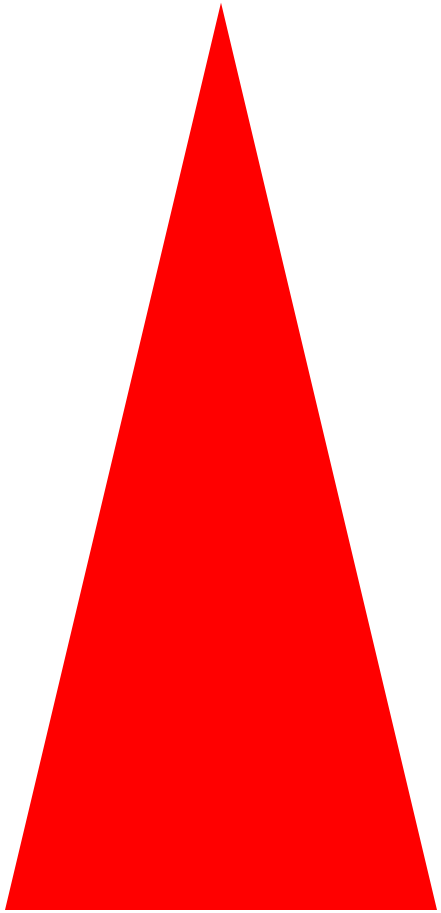
Hierarchy of evidence

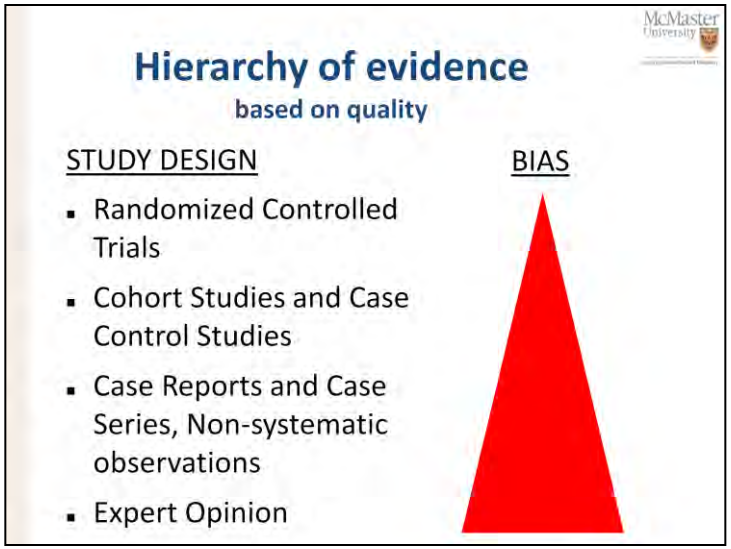
based on quality

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations
- Expert Opinion

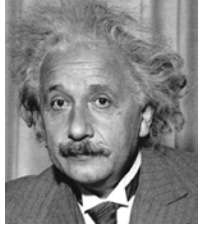
BIAS





The hierarchy of evidence is typically described as in this slide. Randomized control trials are on top, cohort studies and case control studies follow. Case reports and case series non systematic observations are further below and expert opinion is at the very bottom. This has to do with our belief that bias decreases as we move from the bottom of this hierarch to the top of this hierarchy and obviously this is very bad news for experts because their opinion is not valued or is believed to be extremely biased. I will demonstrate on the next slides that this perception or conceptualization of a hierarchy of evidence is likely to be flawed.

"Everything should be made as simple as possible but not simpler."



McMaster University
Inspiring Innovation and Discovery

Explain the following?

- Confounding, effect modification & ext. validity
- Concealment of randomization
- Blinding (who is blinded in a double blinded study?)
- Intention to treat analysis and its correct application
- P-values and confidence intervals

McMaster University
Inspiring Innovation and Discovery

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

HULTON/GETTY

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

BMJ, 2003

Let's take the example from the BMJ Christmas edition in 2003, looking at the use of parachute to prevent death and major trauma related to gravitational challenge, as systematic review of randomized control trials. You can take a guess how many randomized control trials the authors actually identified in their search for evidence. In the context of this particular publication it must be emphasized that it is a BMJ Christmas edition indicating that a topic was addressed in perhaps a not very serious way.

However the authors actually did transmit a very important message. And this important message is very relevant to the way that we look at the quality of evidence or the confidence in an estimate of an effect.

In the GRADE approach one might have looked for the actual observational data that are available to support the use of parachutes in this particular context. And low and behold if we actually look for evidence we would have found evidence that perhaps is better than the evidence in many many health care decision making contexts. There is registry that is maintained by the US Parachute Association and registers every single jump from an airplane. So in 2007, trying to make any decision here evidence based, there were over 2 million jumps that were registered by this organization. And indeed there were 821 injuries and 18 deaths indicating that the use of parachutes is not free of harm but the relative risk reduction calculated on the basis of these events and the total number of jumps was greater than 99.9%. The challenge here is to think of health care interventions that come with an effect that is large enough to make us confident.

McMaster University
Inspiring Innovation and Discovery

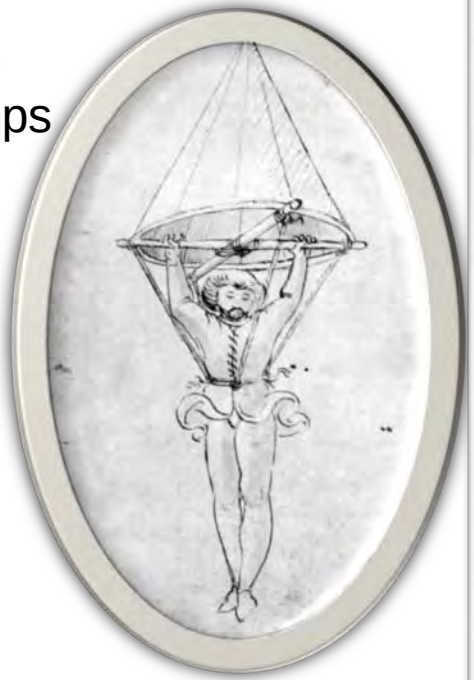
Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

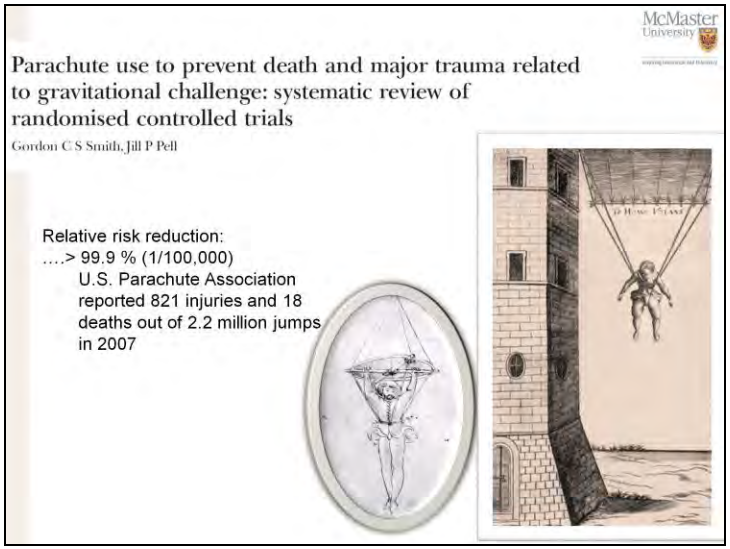
Gordon C S Smith, Jill P Pell

Relative risk reduction:

….> 99.9 % (1/100,000)

U.S. Parachute Association reported 821 injuries and 18 deaths out of 2.2 million jumps in 2007





This magnitude of effect certainly would make us confident that parachutes do in fact work in the way that they were built today to prevent death in the majority of cases. It is not, however the mechanics that were considered by physicists such as Newton or geniuses such as Leonardo da Vinci when thinking about the use of parachutes for the avoidance of gravitational challenges.

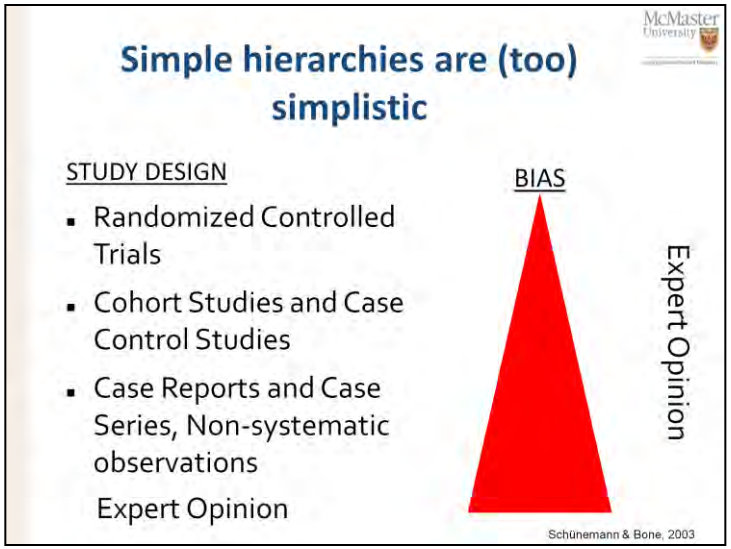
Simple hierarchies are (too) simplistic

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations

Expert Opinion

Schünemann & Bone, 2003



What this indicates is that simply hierarchies are likely too simplistic. Sometimes observational data provide us with very high confidence that in effect exists and in fact, the conduct of randomized control trials would be either unnecessary or ethical. What it also exemplifies is, that expert opinion is required to interpret the available evidence, such as the evidence from observational studies for this particular example.

McMaster University
Inspiring Innovation and Discovery

Why bother about grading?

- People always draw conclusions about:
  - Quality of evidence
  - Strength of recommendation
- Systematic and explicit approaches can help:
  - Protect against errors
  - Resolve disagreements
  - Facilitate critical appraisal
  - Communicate information

Getting from evidence to recommendations - GRADE

Recommendations are judgments:

- Quality of evidence
- Trade off between benefits and harms
- Values and preferences
- Resource use

But judgments need to be based on the best available evidence and transparent

McMaster University
Inspiring Innovation and Discovery

Guidelines and questions

Guidelines are a way of answering questions about clinical, communication, organisational or policy interventions, in the hope of improving health care or health policy.

It is therefore helpful to structure a guideline in terms of answerable questions.

Types of questions

Background Questions

Definition: *What is H5N1 Influenza?*

Mechanism: *What is the mechanism of action of oseltamivir therapy?*

Foreground Questions

Efficacy: *In patients with H5N1 influenza, does oseltamivir improve survival?*

McMaster University
Inspiring Innovation and Discovery

Framing a foreground question

P

I

C

O

McMaster University
Inspiring Innovation and Discovery

Framing a foreground question

Population:

Intervention:

Comparison:

Outcomes:

Case scenario

A 13 year old girl who lives in rural Indonesia presented with flu symptoms and developed severe respiratory distress over the course of the last 2 days. She required intubation. The history reveals that she shares her living quarters with her parents and her three siblings. At night the family's chicken stock shares this room too and several chicken had died unexpectedly a few days before the girl fell sick.

Potential interventions: antivirals, such as neuraminidase inhibitors oseltamivir and zanamivir

McMaster University
Inspiring Innovation and Discovery

What are examples of:

- **Background questions**

- **Foreground questions**
  - **P**opulation:
  - **I**ntervention:
  - **C**omparison:
  - **O**utcomes:

We distinguish different types of questions. Background questions from foreground questions. Background questions for instance deal with definitions, what is contact investigation in TB. Mechanisms, what is the mechanism of transmission of TB, while foreground questions typically deal with questions that lead themselves or lend themselves to recommendations. So for instance, a foreground questions might address the issue of efficacy. What proportion of people who have contact with new or recurrent cases of TB are correctly diagnosed? It is not only efficacy of interventions but also the efficacy of certain diagnostic strategies that could be considered as a typical foreground question. Other examples include the definition of what is avian influenza? What is the mechanism of transmission of the avian influenza virus and the efficacy might relate to what effect do anti-virals have on patient important outcomes such as reducing mortality or reducing hospitalizations. These type of foreground questions once again lend themselves to develop recommendations and guidelines.

There are specific ways of framing a foreground question. The PICO framework is frequently used. It defines the population, the intervention, the comparison and the outcomes. This framework once again is widely used and allows a structured development of a guideline. Take this example from a guideline regarding contact investigation in tuberculosis. A PICO question may read as follows, in people living interventions low and middle income countries who have contact with new or recurrent cases of TB, does contact investigation compare to no contact investigation, reduce overall mortality, reduce consequences of TB infection, cause adverse effects of treatment, how does it increase resource use, or does it increase resource use and so on. This question exemplifies that the population, the intervention, the comparator and the outcomes are clearly defined. One can think of this particular question as also lending itself to the development of sub-questions. The population could be further separated into people in various risk groups and different investigations, contact investigations could be compared against each other.

McMaster University
Inspiring Innovation and Discovery

Framing a foreground question

Population: Avian Flu/influenza A (H5N1) patients

Intervention: Oseltamivir (or Zanamivir)

Comparison: No pharmacological intervention

Outcomes: Mortality, hospitalizations, resource use, adverse outcomes, antimicrobial resistance

Choosing outcomes

- Every decision comes with desirable and undesirable consequences

  →Developing recommendations must include a consideration of desirable and undesirable outcomes

- Outcomes should be patient important outcomes.

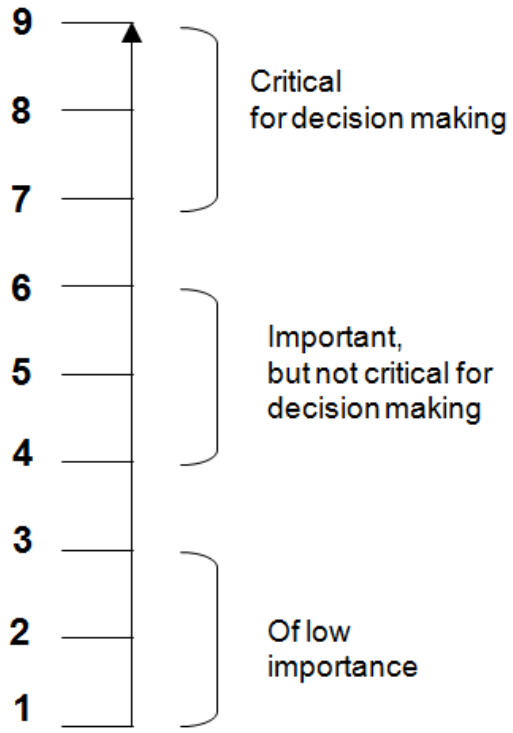
Choosing outcomes

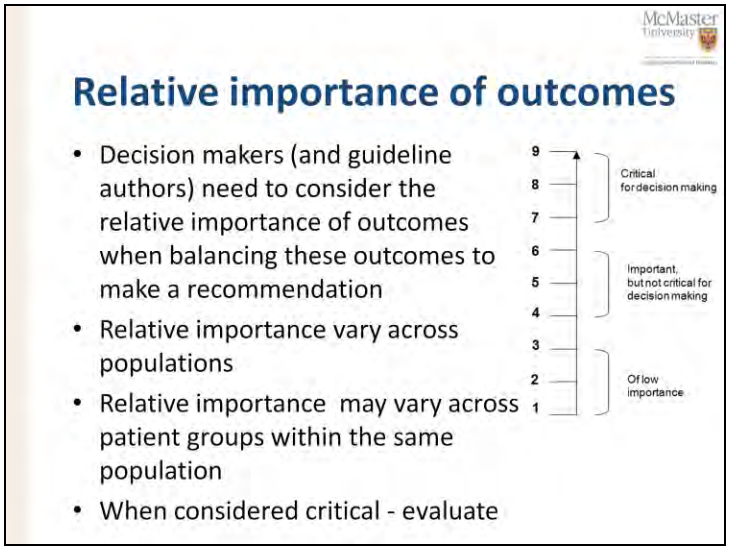
- desirable outcomes
  - lower mortality
  - reduced hospital stay
  - reduced duration of disease
  - reduced resource expenditure

- undesirable outcomes
  - adverse reactions
  - the development of resistance
  - costs of treatment

Relative importance of outcomes

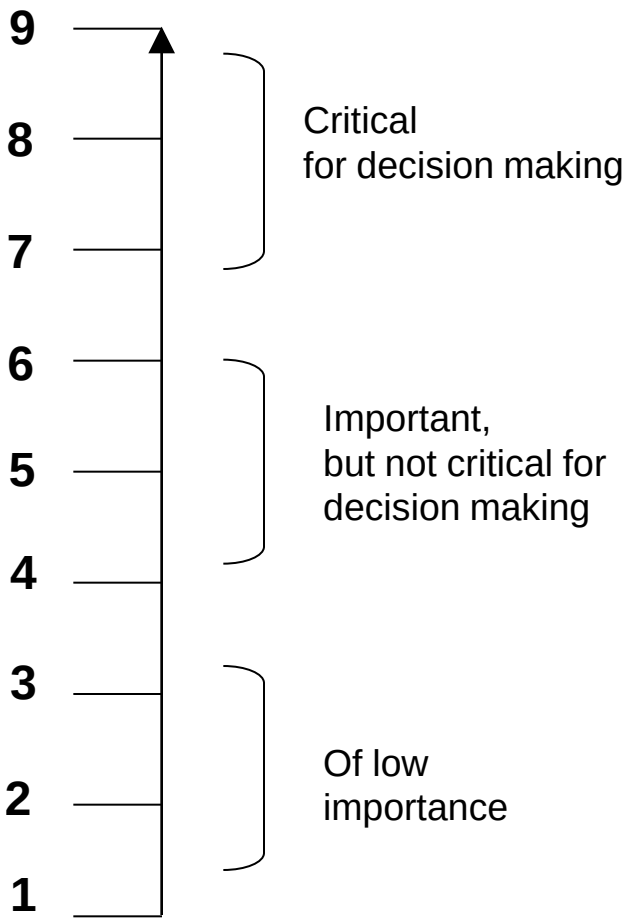
- Decision makers (and guideline authors) need to consider the relative importance of outcomes when balancing these outcomes to make a recommendation
- Relative importance vary across populations
- Relative importance may vary across patient groups within the same population
- When considered critical - evaluate

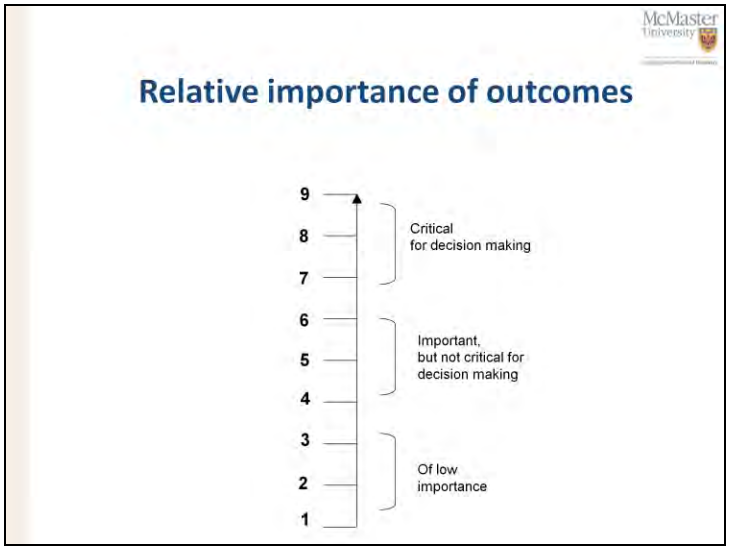




A challenging part of a development of questions is choosing outcomes. We distinguish desirable outcomes such as lower mortality, reducing hospital stay, reducing duration of disease, reduced resource expenditure from undesirable outcomes that basically represent the opposite, such as increase adverse reactions to development of resistance or the cost of treatment. It is important to consider that every decision in life comes with desirable and undesirable consequences and the development of recommendations must include a consideration of these desirable and undesirable consequences. In other words an evaluation of whether net harm when comparing two interventions is avoided.

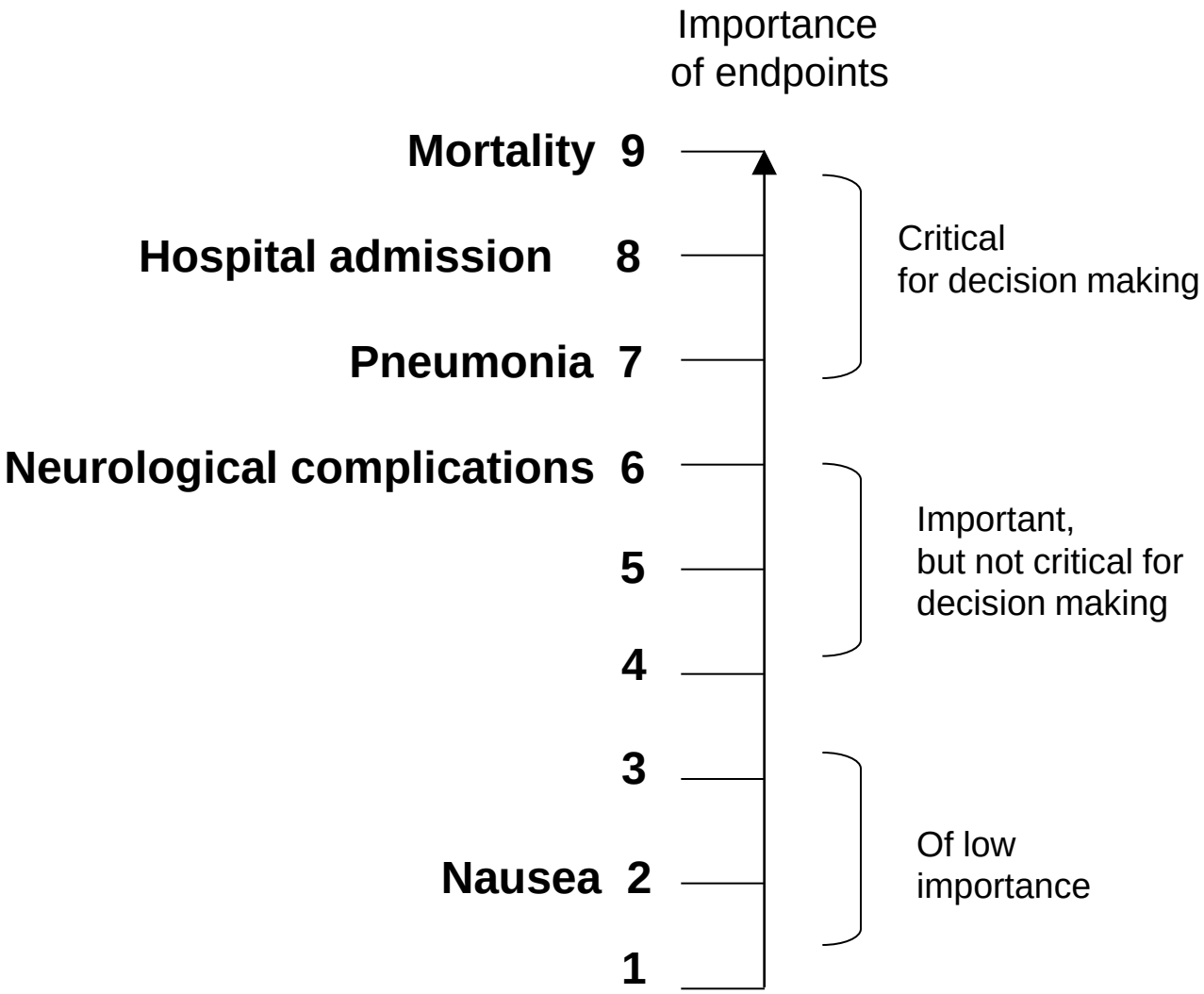
Relative importance of outcomes





One approach to leading decision makers to include the consideration of the relative importance of outcomes is described here. Decision makers and guideline authors need to consider the relative important of outcomes when balancing these outcomes to make a recommendation. Not all outcomes are of similar importance in other words. This relative importance can vary across populations and the relative importance may vary across patient groups within the same population. These are important factors to consider, simple ways of assessing the relative importance of outcomes are the use of scales, such as the scales shown on this slide, distinguishing outcomes that are of low importance, outcomes that may be important but not critical for decision making and those that are critical for decision making. The underlying principal is that when outcomes are considered critical they should be evaluated.

Hierarchy of outcomes according to their importance to assess the effect of oseltamivir in patients with H5N1 influenza



Choosing outcomes

- What if what is important is not measured?
- What if what is measured is not important?
- How do we make sure we've covered all important outcomes?

Choosing outcomes

- Desirable outcomes
  - lower mortality
  - reduced hospital stay
  - reduced duration of disease
  - reduced resource expenditure
- Undesirable outcomes
  - adverse reactions
  - the development of resistance
  - costs of treatment
- Every decision comes with desirable and undesirable consequences

  →Developing recommendations must include a consideration of desirable and undesirable outcomes in terms of the quality of evidence

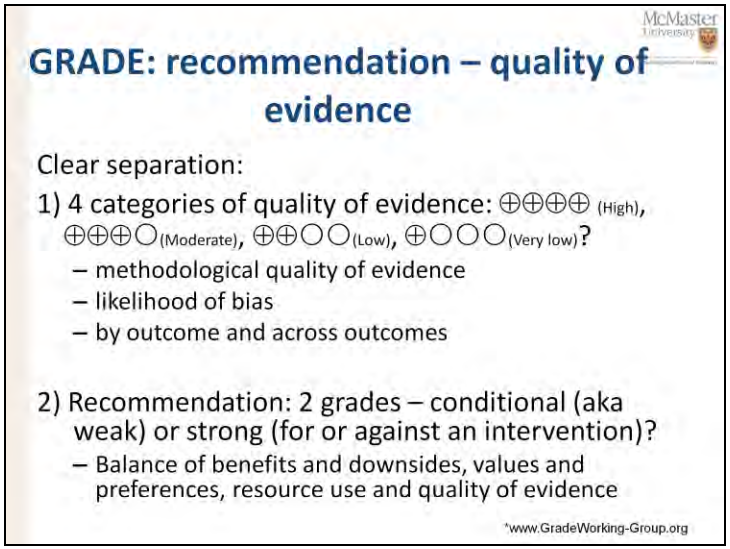
GRADE: recommendation – quality of evidence

Clear separation:

1) 4 categories of quality of evidence: ⨁⨁⨁⨁ (High), ⨁⨁⨁◯ (Moderate), ⨁⨁◯◯ (Low), ⨁◯◯◯ (Very low)?
   - methodological quality of evidence
   - likelihood of bias
   - by outcome and across outcomes

2) Recommendation: 2 grades – conditional (aka weak) or strong (for or against an intervention)?
   - Balance of benefits and downsides, values and preferences, resource use and quality of evidence

*www.GradeWorking-Group.org



You see from that slide that GRADE separates two issues. It separates recommendations from the quality of the evidence. There are 4 categories of the quality of evidence ranging from 4+, also called high to 1+, called very low. The assessment of the quality of evidence that clearly can be considered as a continuum but benefits from an expression in categories for communication purposes is based on the methodological quality of the evidence. In other words, the likelihood of bias, but it is not restricted to internal validity that has been typically considered bias, but it relates to what the possibility of bias is when we think about a health care question and look at the evidence that is available. It includes issues around generalizability or transferability of findings; it includes issues that influence our confidence and estimate of effect that go beyond the risk of bias such as publication bias, inconsistency and impression. This is done by outcome and across outcome and once again this is separated from developing recommendations. There are two Grades of recommendations, they are either conditional, also known as weak or strong and those recommendations are made for or against an intervention. WHO has typically preferred in its terminology the word conditional, but weak is a synonymous term that can be used. And once again the strength of a recommendation depends on the balance of benefits and downsides, values and preferences, resource use and the quality of evidence.

McMaster University
Inspiring Innovation and Discovery

GRADE Quality of Evidence

In the context of making recommendations:

- The quality of evidence reflects the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.

Figure 1. *Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street Journal).*

Likelihood of and confidence in an outcome

Figure 1. *Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street Journal).*

We can look at this as depicted in this cartoon. The likelihood of and the confidence in an outcome. In the cartoon one meteorologist is saying to another, I figure there is a 40% chance of showers and a 10% chance we know what we are talking about. Once again, this expresses our confidence in an estimate of effect and the likelihood that it actually occurs. For instance, the confidence intervals around the 404 chance of showers estimate may be very tight. They may in fact be based on modeling that has come up with confidence intervals that range from 35 – 45 %. However, the development of the model or the application of the model from one setting to another may leave us with very little confidence that the estimate is actually correct for the particular setting. Just imagine that model being developed in Australia and applied to North America. Once again, this is similar to how we look at the confidence in evidence in the GRADE approach.

Determinants of quality

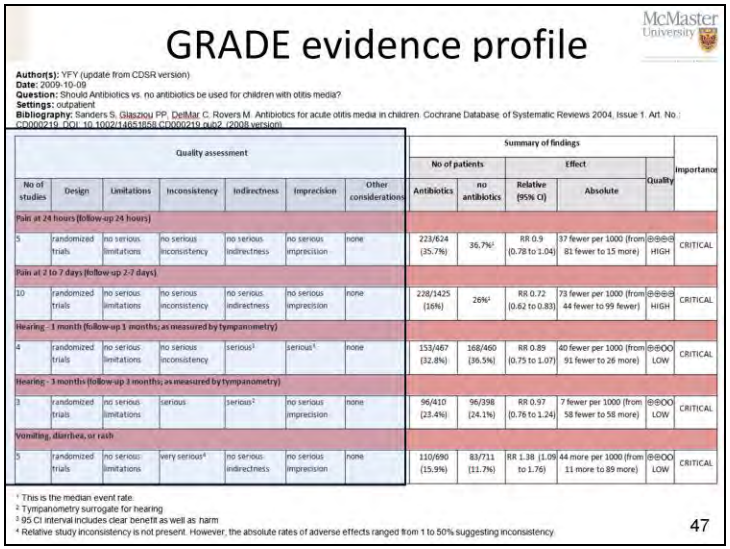
- RCTs ⊕⊕⊕⊕
- observational studies ⊕⊕◯◯
- 5 factors that can lower quality
  1. limitations in detailed design and execution *(risk of bias criteria)*
  2. Inconsistency *(or heterogeneity)*
  3. Indirectness *(PICO and applicability)*
  4. Imprecision *(number of events and confidence intervals)*
  5. Publication bias
- 3 factors can increase quality
  1. large magnitude of effect
  2. all plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
  3. dose-response gradient

GRADE evidence profile

Author(s): YFY (update from CDSR version)
Date: 2009-10-09
Question: Should Antibiotics vs. no antibiotics be used for children with otitis media?
Settings: outpatient
Bibliography: Sanders S, Glasziou PP, DelMar C, Rovers M. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000219. DOI: 10.1002/14651858.CD000219.pub2. (2008 version)

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Antibiotics	no antibiotics	Relative (95% CI)	Absolute		
Pain at 24 hours (follow-up 24 hours)												
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	223/624 (35.7%)	36.7%[1]	RR 0.9 (0.78 to 1.04)	37 fewer per 1000 (from 81 fewer to 15 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Pain at 2 to 7 days (follow-up 2-7 days)												
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	228/1425 (16%)	26%[1]	RR 0.72 (0.62 to 0.83)	73 fewer per 1000 (from 44 fewer to 99 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)												
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none	153/467 (32.8%)	168/460 (36.5%)	RR 0.89 (0.75 to 1.07)	40 fewer per 1000 (from 91 fewer to 26 more)	⊕⊕OO LOW	CRITICAL
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)												
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none	96/410 (23.4%)	96/398 (24.1%)	RR 0.97 (0.76 to 1.24)	7 fewer per 1000 (from 58 fewer to 58 more)	⊕⊕OO LOW	CRITICAL
Vomiting, diarrhea, or rash												
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none	110/690 (15.9%)	83/711 (11.7%)	RR 1.38 (1.09 to 1.76)	44 more per 1000 (from 11 more to 89 more)	⊕⊕OO LOW	CRITICAL

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged from 1 to 50% suggesting inconsistency.



GRADE evidence profile

McMaster University

Author(s): YFY (update from CDSR version)
Date: 2009-10-09
Question: Should Antibiotics vs. no antibiotics be used for children with otitis media?
Settings: outpatient
Bibliography: Sanders S, Glasziou PP, DelMar C, Rovers M. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000219. DOI: 10.1002/14651858.CD000219.pub2. (2008 version)

Quality assessment							Summary of findings					
							No of patients		Effect			Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Antibiotics	no antibiotics	Relative (95% CI)	Absolute	Quality	
Pain at 24 hours (follow-up 24 hours)												
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	223/624 (35.7%)	36.7%[1]	RR 0.9 (0.78 to 1.04)	37 fewer per 1000 (from 81 fewer to 15 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Pain at 2 to 7 days (follow-up 2-7 days)												
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	228/1425 (16%)	26%[2]	RR 0.72 (0.62 to 0.83)	73 fewer per 1000 (from 44 fewer to 99 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)												
4	randomized trials	no serious limitations	no serious inconsistency	serious[3]	serious[4]	none	153/467 (32.8%)	168/460 (36.5%)	RR 0.89 (0.75 to 1.07)	40 fewer per 1000 (from 91 fewer to 26 more)	⊕⊕OO LOW	CRITICAL
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)												
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none	96/410 (23.4%)	96/398 (24.1%)	RR 0.97 (0.78 to 1.24)	7 fewer per 1000 (from 58 fewer to 58 more)	⊕⊕OO LOW	CRITICAL
Vomiting, diarrhea, or rash												
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none	110/690 (15.9%)	83/711 (11.7%)	RR 1.38 (1.09 to 1.76)	44 more per 1000 (from 11 more to 89 more)	⊕⊕OO LOW	CRITICAL

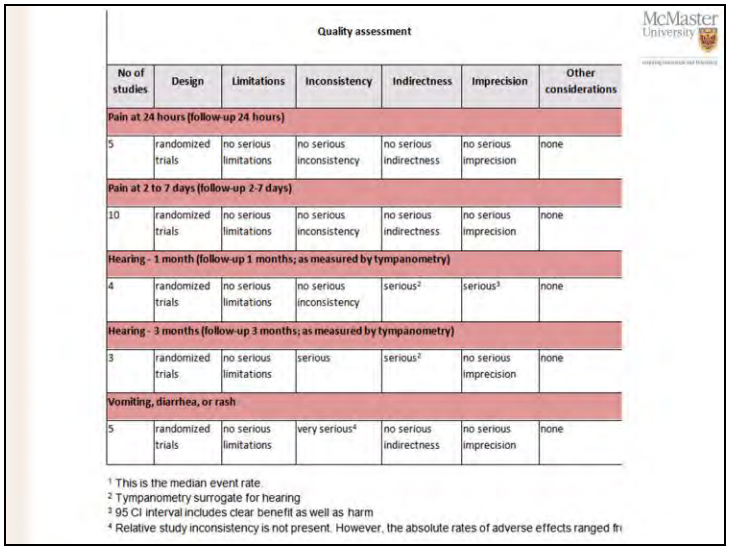
[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged from 1 to 50% suggesting inconsistency.

47

GRADE evidence syntheses describe a summary of the key results from a systematic review that guideline panel members can use to produce recommendations in clinical practice guidelines or other health care guidelines. We typically describe the GRADE evidence syntheses as evidence profiles or Summary of Findings tables. They present the quality of the evidence or the confidence in the estimate of an effect for a related outcome based a body of evidence, they present the magnitude of an effect typically both in relative and absolute terms both for dichotomous as well as continuous outcomes and they provide a transparent description of judgments about the evidence or provide further explanation about other important aspects of an evidence synthesis.

Quality assessment						
No of studies	**Design**	**Limitations**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**
Pain at 24 hours (follow-up 24 hours)						
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none
Pain at 2 to 7 days (follow-up 2-7 days)						
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)						
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)						
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none
Vomiting, diarrhea, or rash						
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged fr



Quality assessment

No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations
Pain at 24 hours (follow-up 24 hours)						
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none
Pain at 2 to 7 days (follow-up 2-7 days)						
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)						
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)						
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none
Vomiting, diarrhea, or rash						
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged fr

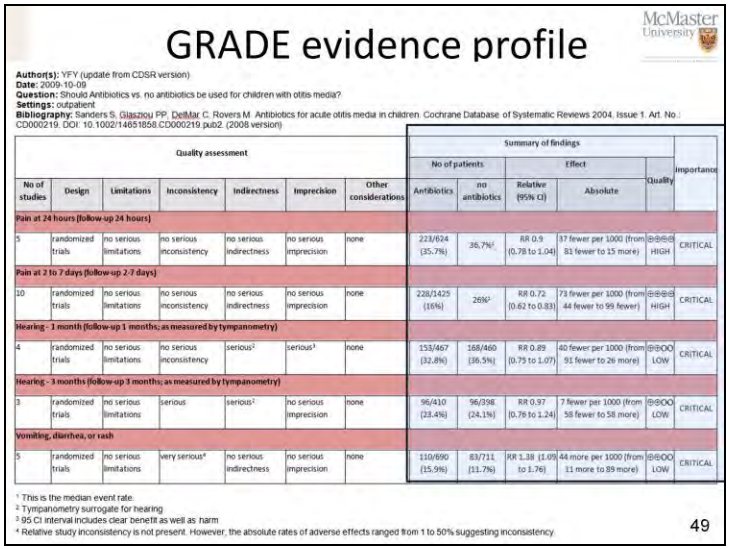
This shows the left side of the table in greater detail.

GRADE evidence profile

Author(s): YFY (update from CDSR version)
Date: 2009-10-09
Question: Should Antibiotics vs. no antibiotics be used for children with otitis media?
Settings: outpatient
Bibliography: Sanders S, Glasziou PP, DelMar C, Rovers M. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000219. DOI: 10.1002/14651858.CD000219.pub2. (2008 version)

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Antibiotics	no antibiotics	Relative (95% CI)	Absolute		
Pain at 24 hours (follow-up 24 hours)												
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	223/624 (35.7%)	36.7%[1]	RR 0.9 (0.78 to 1.04)	37 fewer per 1000 (from 81 fewer to 15 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Pain at 2 to 7 days (follow-up 2-7 days)												
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	228/1425 (16%)	26%[1]	RR 0.72 (0.62 to 0.83)	73 fewer per 1000 (from 44 fewer to 99 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)												
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none	153/467 (32.8%)	168/460 (36.5%)	RR 0.89 (0.75 to 1.07)	40 fewer per 1000 (from 91 fewer to 26 more)	⊕⊕OO LOW	CRITICAL
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)												
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none	96/410 (23.4%)	96/398 (24.1%)	RR 0.97 (0.76 to 1.24)	7 fewer per 1000 (from 58 fewer to 58 more)	⊕⊕OO LOW	CRITICAL
Vomiting, diarrhea, or rash												
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none	110/690 (15.9%)	83/711 (11.7%)	RR 1.38 (1.09 to 1.76)	44 more per 1000 (from 11 more to 89 more)	⊕⊕OO LOW	CRITICAL

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged from 1 to 50% suggesting inconsistency.



GRADE evidence profile

McMaster University

Author(s): YFY (update from CDSR version)
Date: 2009-10-09
Question: Should Antibiotics vs. no antibiotics be used for children with otitis media?
Settings: outpatient
Bibliography: Sanders S, Glasziou PP, DelMar C, Rovers M. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000219. DOI: 10.1002/14651858.CD000219.pub2. (2008 version)

Quality assessment							Summary of findings				
							No of patients		Effect		Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Antibiotics	no antibiotics	Relative (95% CI)	Absolute	Quality	
Pain at 24 hours (follow-up 24 hours)												
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	223/624 (35.7%)	36.7%[1]	RR 0.9 (0.78 to 1.04)	37 fewer per 1000 (from 81 fewer to 15 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Pain at 2 to 7 days (follow-up 2-7 days)												
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	228/1425 (16%)	26%[1]	RR 0.72 (0.62 to 0.83)	73 fewer per 1000 (from 44 fewer to 99 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)												
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none	153/467 (32.8%)	168/460 (36.5%)	RR 0.89 (0.75 to 1.07)	40 fewer per 1000 (from 91 fewer to 26 more)	⊕⊕OO LOW	CRITICAL
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)												
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none	96/410 (23.4%)	96/398 (24.1%)	RR 0.97 (0.78 to 1.24)	7 fewer per 1000 (from 58 fewer to 58 more)	⊕⊕OO LOW	CRITICAL
Vomiting, diarrhea, or rash												
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none	110/690 (15.9%)	83/711 (11.7%)	RR 1.38 (1.09 to 1.76)	44 more per 1000 (from 11 more to 89 more)	⊕⊕OO LOW	CRITICAL

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged from 1 to 50% suggesting inconsistency.

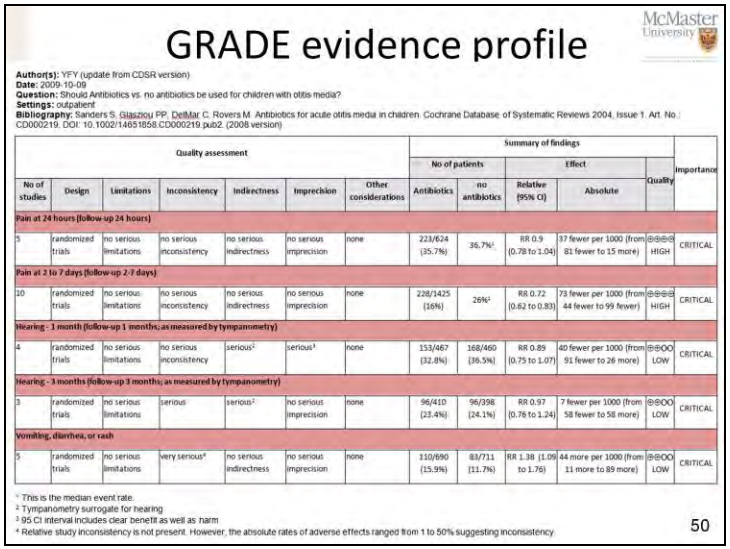
This shows the rightside of the table in greater detail.

GRADE evidence profile

Author(s): YFY (update from CDSR version)
Date: 2009-10-09
Question: Should Antibiotics vs. no antibiotics be used for children with otitis media?
Settings: outpatient
Bibliography: Sanders S, Glasziou PP, DelMar C, Rovers M. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000219. DOI: 10.1002/14651858.CD000219.pub2. (2008 version)

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Antibiotics	no antibiotics	Relative (95% CI)	Absolute		
Pain at 24 hours (follow-up 24 hours)												
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	223/624 (35.7%)	36.7%[1]	RR 0.9 (0.78 to 1.04)	37 fewer per 1000 (from 81 fewer to 15 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Pain at 2 to 7 days (follow-up 2-7 days)												
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	228/1425 (16%)	26%[1]	RR 0.72 (0.62 to 0.83)	73 fewer per 1000 (from 44 fewer to 99 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)												
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none	153/467 (32.8%)	168/460 (36.5%)	RR 0.89 (0.75 to 1.07)	40 fewer per 1000 (from 91 fewer to 26 more)	⊕⊕OO LOW	CRITICAL
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)												
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none	96/410 (23.4%)	96/398 (24.1%)	RR 0.97 (0.76 to 1.24)	7 fewer per 1000 (from 58 fewer to 58 more)	⊕⊕OO LOW	CRITICAL
Vomiting, diarrhea, or rash												
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none	110/690 (15.9%)	83/711 (11.7%)	RR 1.38 (1.09 to 1.76)	44 more per 1000 (from 11 more to 89 more)	⊕⊕OO LOW	CRITICAL

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged from 1 to 50% suggesting inconsistency.



GRADE evidence profile

McMaster University

Author(s): YFY (update from CDSR version)
Date: 2009-10-09
Question: Should Antibiotics vs. no antibiotics be used for children with otitis media?
Settings: outpatient
Bibliography: Sanders S, Glasziou PP, DelMar C, Rovers M. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews 2004, Issue 1. Art. No.: CD000219. DOI: 10.1002/14651858.CD000219.pub2. (2008 version)

Quality assessment							Summary of findings				
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Antibiotics	no antibiotics	Relative (95% CI)	Absolute		
Pain at 24 hours (follow-up 24 hours)												
5	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	223/624 (35.7%)	36.7%[1]	RR 0.9 (0.78 to 1.04)	37 fewer per 1000 (from 81 fewer to 15 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Pain at 2 to 7 days (follow-up 2-7 days)												
10	randomized trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	228/1425 (16%)	26%[1]	RR 0.72 (0.62 to 0.83)	73 fewer per 1000 (from 44 fewer to 99 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Hearing - 1 month (follow-up 1 months; as measured by tympanometry)												
4	randomized trials	no serious limitations	no serious inconsistency	serious[2]	serious[3]	none	153/467 (32.8%)	168/460 (36.5%)	RR 0.89 (0.75 to 1.07)	40 fewer per 1000 (from 91 fewer to 26 more)	⊕⊕OO LOW	CRITICAL
Hearing - 3 months (follow-up 3 months; as measured by tympanometry)												
3	randomized trials	no serious limitations	serious	serious[2]	no serious imprecision	none	96/410 (23.4%)	96/398 (24.1%)	RR 0.97 (0.78 to 1.24)	7 fewer per 1000 (from 58 fewer to 58 more)	⊕⊕OO LOW	CRITICAL
Vomiting, diarrhea, or rash												
5	randomized trials	no serious limitations	very serious[4]	no serious indirectness	no serious imprecision	none	110/690 (15.9%)	83/711 (11.7%)	RR 1.38 (1.09 to 1.76)	44 more per 1000 (from 11 more to 89 more)	⊕⊕OO LOW	CRITICAL

[1] This is the median event rate.
[2] Tympanometry surrogate for hearing
[3] 95 CI interval includes clear benefit as well as harm
[4] Relative study inconsistency is not present. However, the absolute rates of adverse effects ranged from 1 to 50% suggesting inconsistency.

50

This shows the detailed profile. The judgments that are made about the rating of the evidence, as well as other information are described in footnotes. This presents a second alternative format of the evidence profile. In this case, the question is whether also time of year should be used compared to no antiviral treatment for patients with influenza. In this case observational studies were summarized; once again this is an alternative format where there is again a quality assessment by outcome as well as a summary of findings. In this case you will see that the overall quality is mentioned earlier in this row, the columns that are currently provided here can be replaced if other factors apply, such as publication bias may be replaced with factors about upgrading and the Summary of Findings table again presents information for both relative as well as absolute effects for various baseline risks. The important aspect here and to highlight is the judgments about the quality of evidence are described in these footnotes that are provided there and highlighted in orange.

Strength of recommendation

"The strength of a recommendation reflects the extent to which we can, across the range of patients for whom the recommendations are intended, be confident that desirable effects of a management strategy outweigh undesirable effects."

- Strong or conditional

McMaster University
Inspiring Innovation and Discovery

Implications of a *strong/category A* recommendation

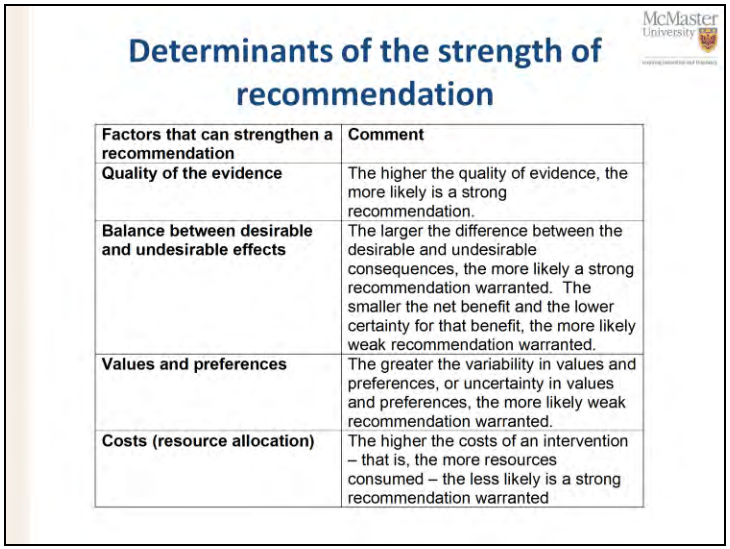
- Patients: Most people in this situation would want the recommended course of action and only a small proportion would not
- Clinicians: Most patients should receive the recommended course of action
- Policy makers: The recommendation can be adapted as a policy in most situations

Implications of a *conditional/weak/category B* recommendation

- Patients: The majority of people in this situation would want the recommended course of action, but many would not
- Clinicians: Be more prepared to help patients to make a decision that is consistent with their own values/decision aids and shared decision making
- Policy makers: There is a need for substantial debate and involvement of stakeholders

Determinants of the strength of recommendation

Factors that can strengthen a recommendation	Comment
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation.
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely weak recommendation warranted.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.
Costs (resource allocation)	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted



This slide shows the four factors that determine the strength and direction of a recommendation. The first is the quality of the evidence and the higher the quality of the evidence the more likely is a strong recommendation. The second is the balance between the benefits and harms; the larger the difference between the benefits and harms there more likely is a strong recommendation warranted. The smaller the net benefit and the lower the certainty for that benefit, the more likely is a weak recommendation warranted. The third is values and preferences. The greater the variability in values and preferences or uncertainty in values and preferences the more likely is a weak or conditional recommendation warranted. And the fourth is cost of resources; the higher the cost for a certain intervention and perhaps the more opportunity costs that the intervention causes, the less likely is a strong recommendation warranted.

ACIP principles

- focus on transparency
- use of evidence of varying strengths consideration of both individual and community health
- adoption or adaptation of an existing evidence-based system
- need for continuous improvement of the process

Agenda

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

Agenda

McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

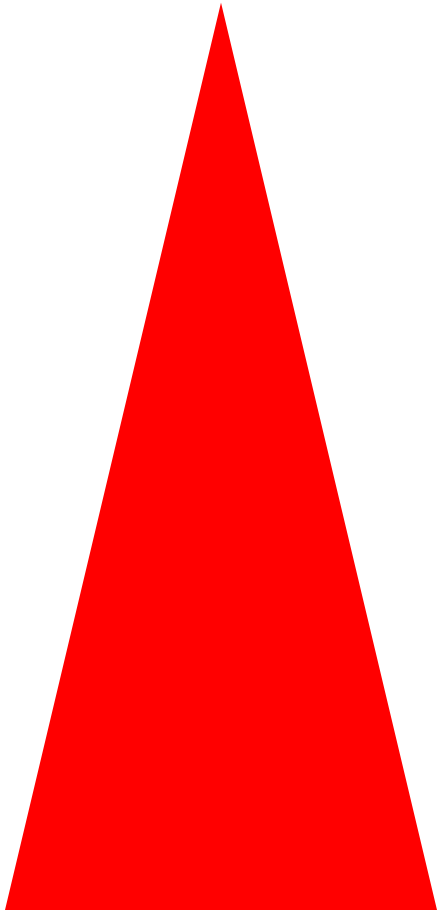
Hierarchy of evidence

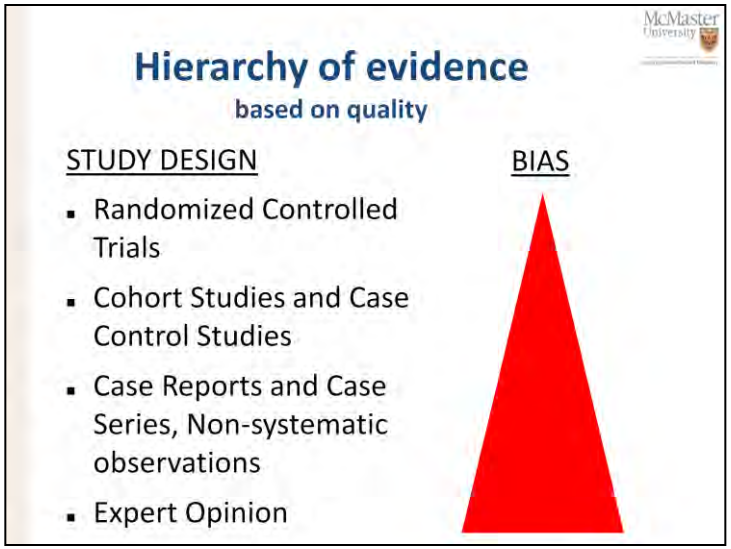
based on quality

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations
- Expert Opinion

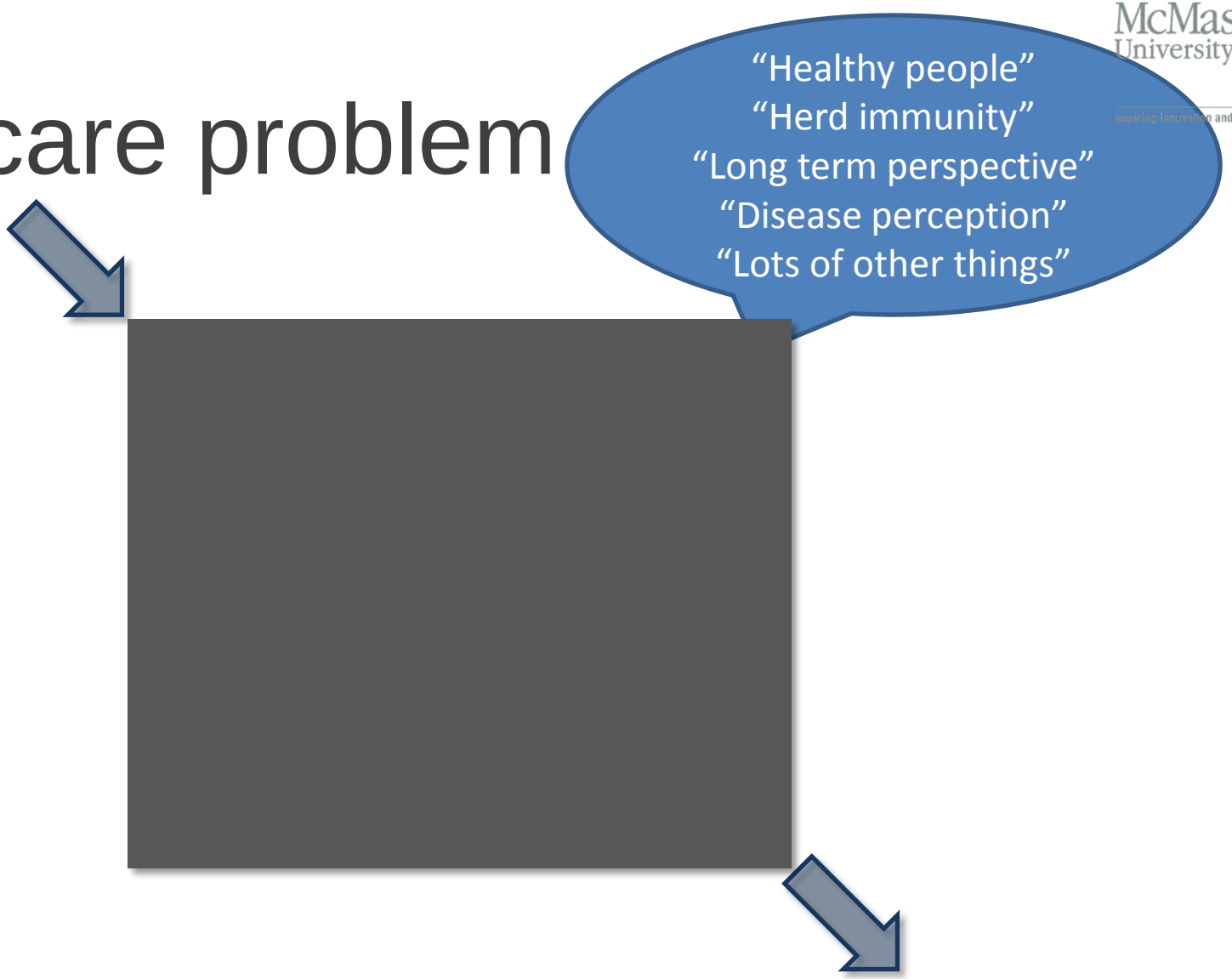
BIAS



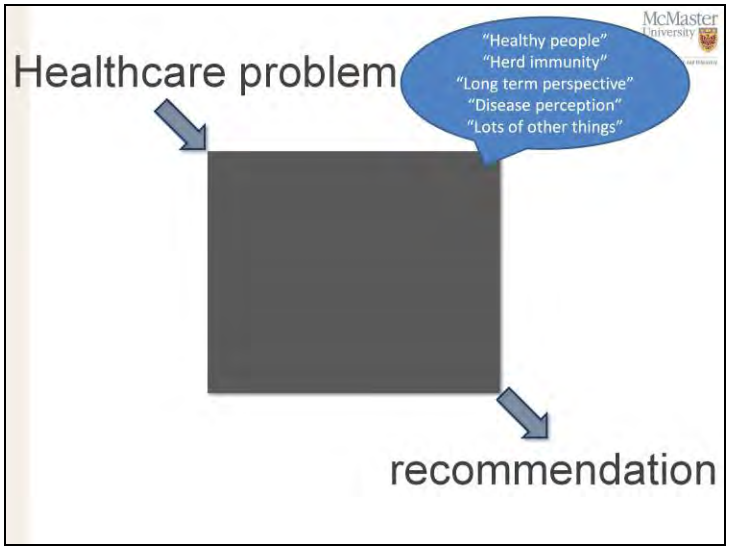


Remember this slide.
What this indicates is that simply hierarchies are likely too simplistic. Sometimes observational data provide us with very high confidence that in effect exists and in fact, the conduct of randomized control trials would be either unnecessary or ethical. What it also exemplifies is, that expert opinion is required to interpret the available evidence, such as the evidence from observational studies for this particular example.

Healthcare problem



recommendation



The process of evaluating the quality has been a black box.

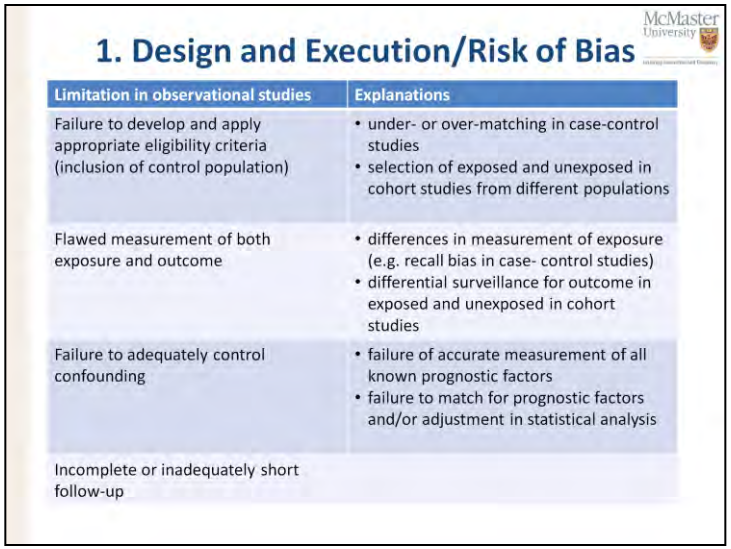
Determinants of quality

- RCTs ⊕⊕⊕⊕
- observational studies ⊕⊕◯◯
- 5 factors that can lower quality
  1. limitations in detailed design and execution *(risk of bias criteria)*
  2. Inconsistency *(or heterogeneity)*
  3. Indirectness *(PICO and applicability)*
  4. Imprecision *(number of events and confidence intervals)*
  5. Publication bias
- 3 factors can increase quality
  1. large magnitude of effect
  2. all plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
  3. dose-response gradient

1. Design and Execution/Risk of Bias

McMaster University
Inspiring Innovation and Discovery

Limitation in observational studies	Explanations
Failure to develop and apply appropriate eligibility criteria (inclusion of control population)	• under- or over-matching in case-control studies • selection of exposed and unexposed in cohort studies from different populations
Flawed measurement of both exposure and outcome	• differences in measurement of exposure (e.g. recall bias in case- control studies) • differential surveillance for outcome in exposed and unexposed in cohort studies
Failure to adequately control confounding	• failure of accurate measurement of all known prognostic factors • failure to match for prognostic factors and/or adjustment in statistical analysis
Incomplete or inadequately short follow-up	



These are the factors to be considered generally when looking at risk of bias in observational studies. Let us begin with an explanation of the criterion of detailed design and execution or risk of bias as a quality criterion. Examples for that are inappropriate selection of exposed and unexposed groups, the failure to adequately measure or control for confounding, selective outcome reporting, failure to blind, for instance outcome assessors which applies both to randomized control studies as well as observational studies, a high loss to follow up, lack of concealment in randomized control trials or a violation of the intention to treat principal when it should not be violated.

McMaster University
Inspiring Innovation and Discovery

1. Design and Execution/Risk of Bias

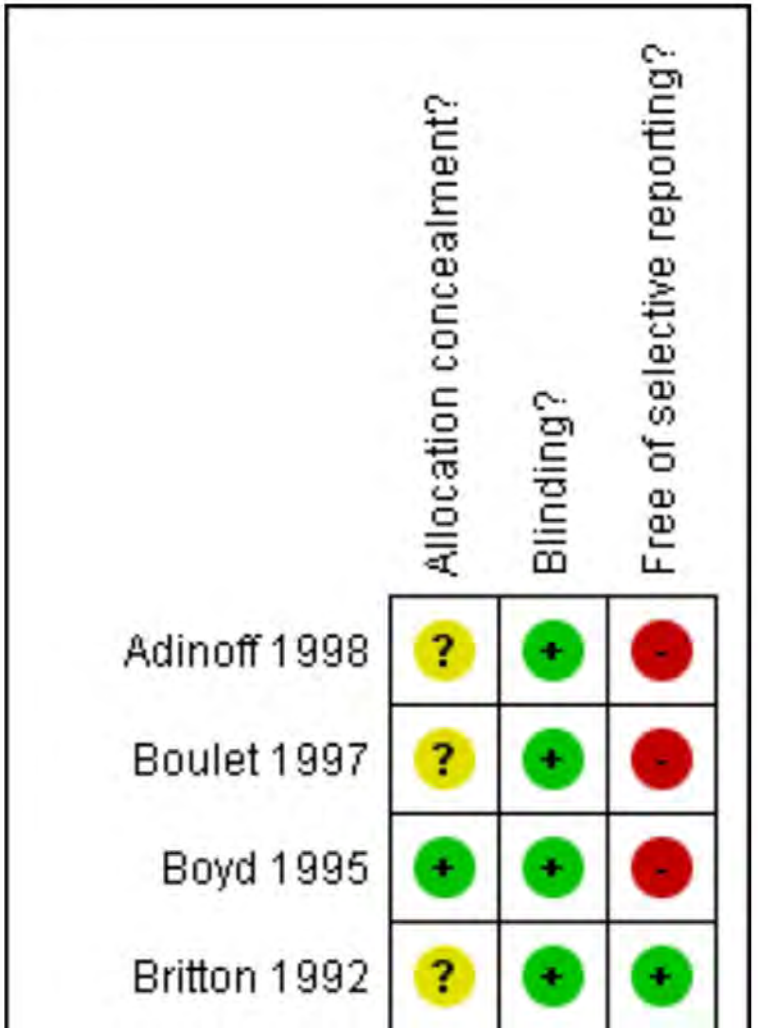
Limitations in RCTs
lack of concealment
intention to treat principle violated
inadequate blinding
loss to follow-up
early stopping for benefit
selective outcome reporting

Design and Execution/RoB

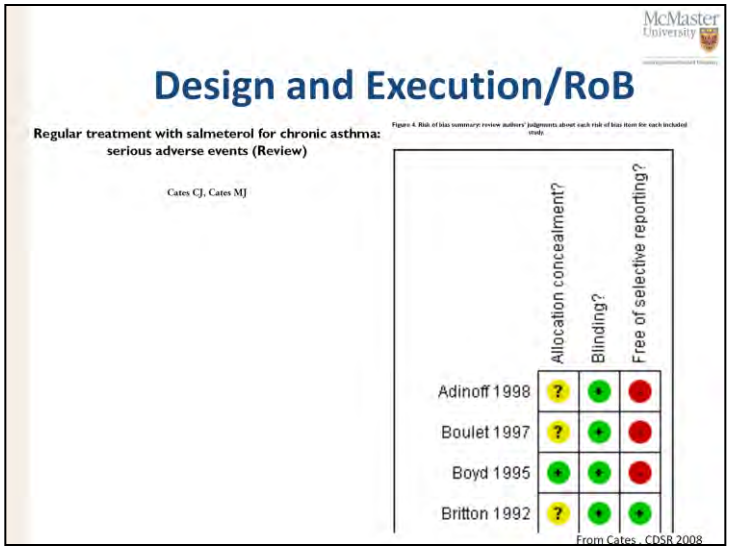
Regular treatment with salmeterol for chronic asthma: serious adverse events (Review)

Cates CJ, Cates MJ

Figure 4. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.



From Cates , CDSR 2008

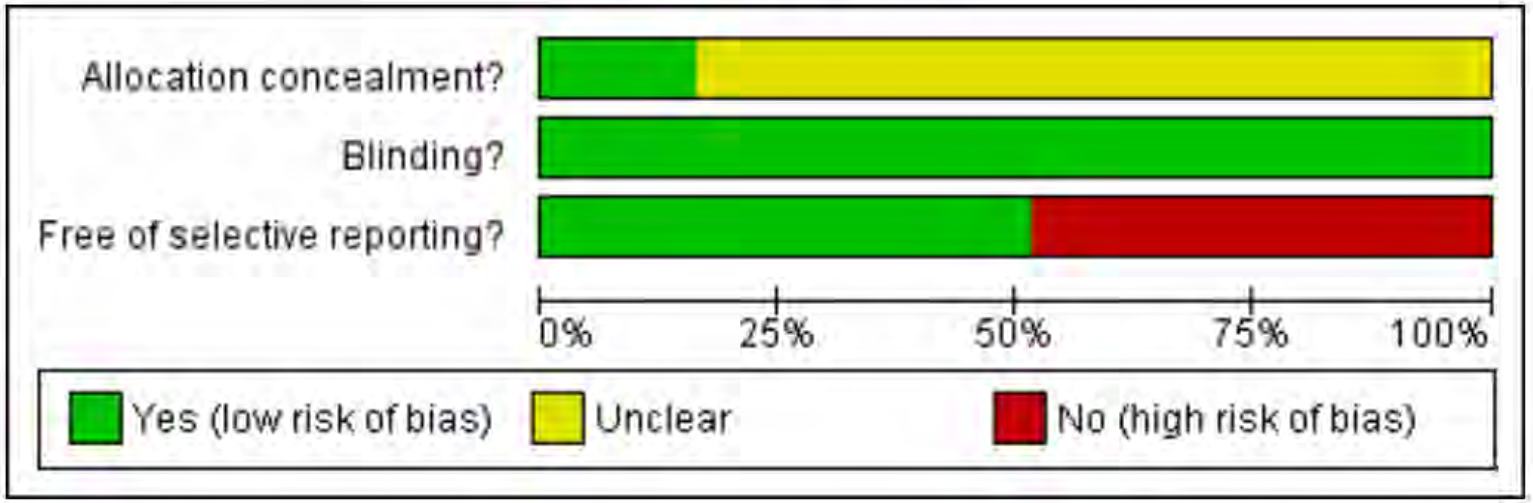


Let's just consider this example. This is an example from a systematic review conducted for the Cochrane Collaboration where the authors were interested in identifying the evidence around serious adverse events related to a particular intervention for chronic asthma. In this case, once again, the outcome of interest is serious adverse events. The authors identified approximately 30 randomized control trials addressing this particular issue. They looked at three particular quality criteria; allocation concealment, blinding and selective outcome reporting. In other words, whether data on serious adverse events were truly reported when the investigators should have had them. As you can tell from this slide, approximately half of the studies did not report on the outcome serious adverse events when they actually had the data available. For instance, many of these studies were submitted for regulatory purposes and serious adverse events must be recorded indicated by the red dots in the column of free of selective reporting. All of these studies were appropriately blinded as per the judgment of the systematic reviewers indicated by green dots, and many of the studies did not provide the information to appropriately assess allocation concealment. What the slides demonstrate is that a detailed assessment of the individual studies is necessary but also that an overall judgment about the underlying body of evidence is required. For instance, if the investigators had found that there is a relative risk that is increased for serious adverse events with this particular medication, even the magnitude of the effect would have been uncertain given that many studies did not report on serious adverse events when they should have reported on them. That means that the true risk could have been larger or smaller.

McMaster University
Inspiring Innovation and Discovery

Design and Execution/RoB

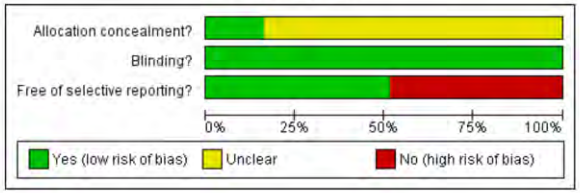
Figure 3. Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.



Overall judgment required

Design and Execution/RoB

Figure 3. Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.



Overall judgment required

This is an alternative way of showing the risk of bias across studies.

Who believes the risk of bias is of concern?

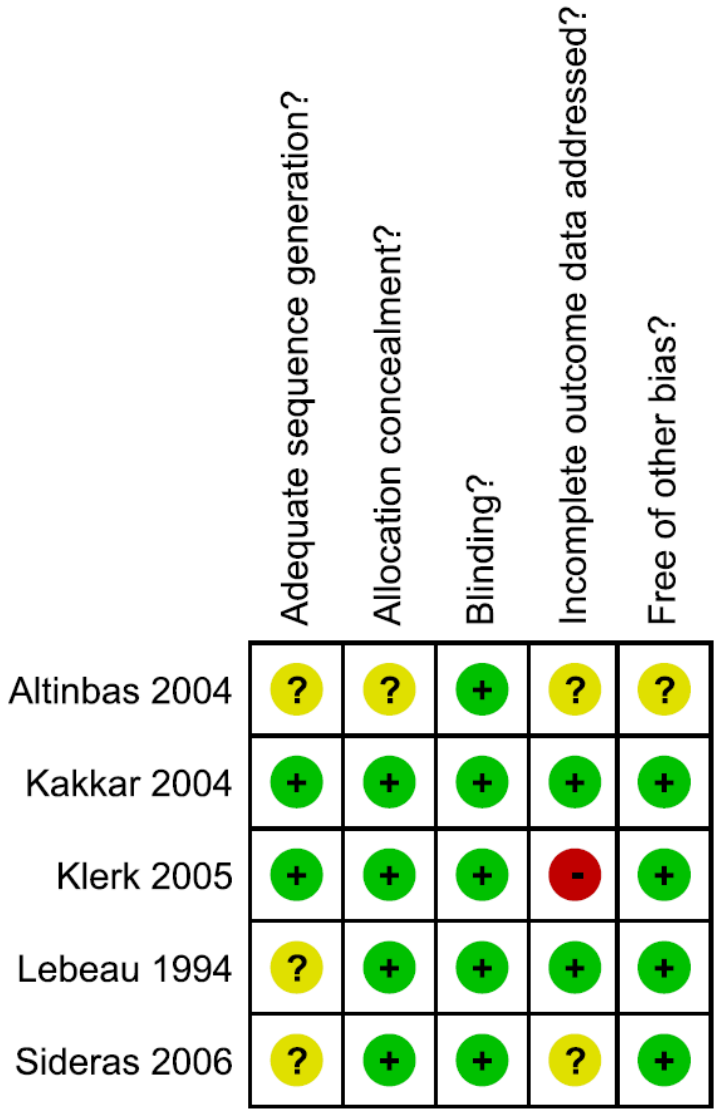
Yes

No

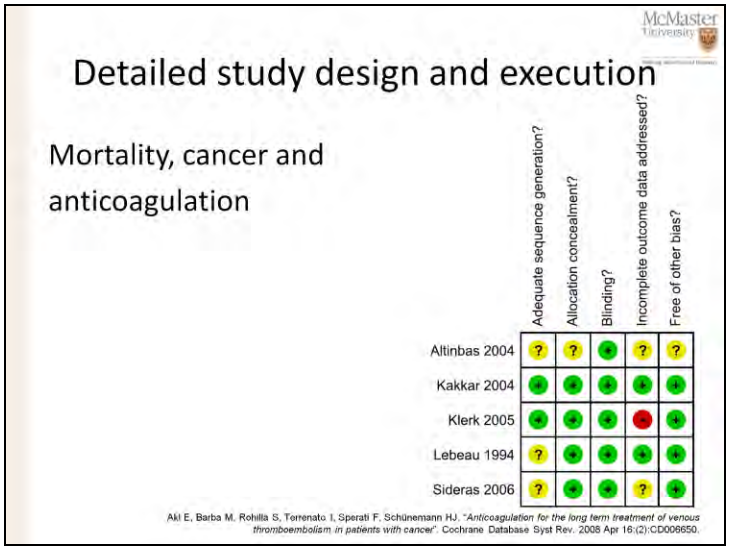
Don’t know or undecided

Detailed study design and execution

Mortality, cancer and anticoagulation



Akl E, Barba M, Rohilla S, Terrenato I, Sperati F, Schünemann HJ. "*Anticoagulation for the long term treatment of venous thromboembolism in patients with cancer*". Cochrane Database Syst Rev. 2008 Apr 16;(2):CD006650.



Now look at this example showing the risk of bias table from a randomized control trial that assessed whether anticoagulation reduces the risk of mortality in patients cancer. There are five randomized control trials and you see the risk of bias assessment here where there is only one significant or important concern about the incomplete outcome data from assessment in the trial by Klerk and colleagues. One might need additional information to make the judgment about whether the risk of bias is important enough to downgrade the quality. One of the pieces of information that one might require is how large or how important this trial is in the overall estimate of effect.

Five trials

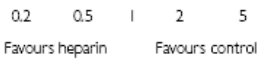
Analysis 01.01. Comparison 01 Heparin vs placebo, Outcome 01 Mortality over duration of study

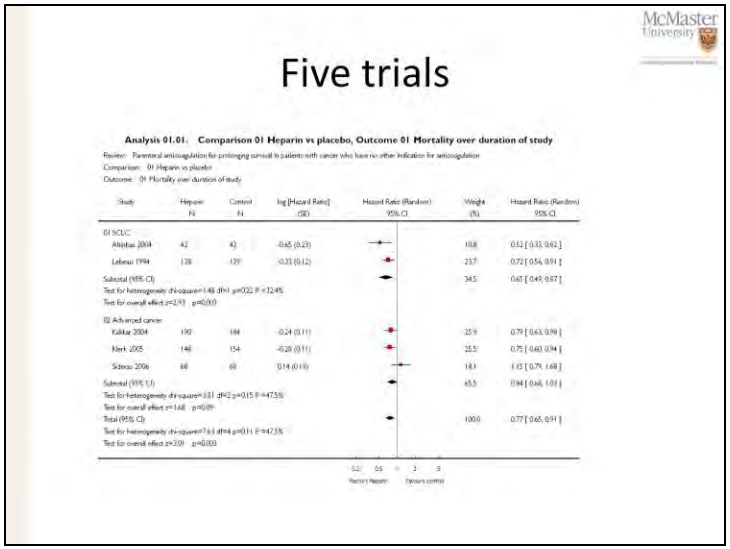
Review: Parenteral anticoagulation for prolonging survival in patients with cancer who have no other indication for anticoagulation

Comparison: 01 Heparin vs placebo

Outcome: 01 Mortality over duration of study

Study	Heparin N	Control N	log [Hazard Ratio] (SE)	Hazard Ratio (Random) 95% CI	Weight (%)	Hazard Ratio (Random) 95% CI
01 SCLC						
Altinbas 2004	42	42	-0.65 (0.23)		10.8	0.52 [0.33, 0.82]
Lebeau 1994	138	139	-0.33 (0.12)		23.7	0.72 [0.56, 0.91]
Subtotal (95% CI)					34.5	0.65 [0.49, 0.87]
Test for heterogeneity chi-square=1.48 df=1 p=0.22 I^2 =32.4%						
Test for overall effect z=2.93 p=0.003						
02 Advanced cancer						
Kakkar 2004	190	184	-0.24 (0.11)		25.9	0.79 [0.63, 0.98]
Klerk 2005	148	154	-0.28 (0.11)		25.5	0.75 [0.60, 0.94]
Sideras 2006	68	69	0.14 (0.19)		14.1	1.15 [0.79, 1.68]
Subtotal (95% CI)					65.5	0.84 [0.68, 1.03]
Test for heterogeneity chi-square=3.81 df=2 p=0.15 I^2 =47.5%						
Test for overall effect z=1.68 p=0.09						
Total (95% CI)					100.0	0.77 [0.65, 0.91]
Test for heterogeneity chi-square=7.63 df=4 p=0.11 I^2 =47.5%						
Test for overall effect z=3.01 p=0.003						





One way of addressing this issue is to look at the forest plot related to the meta-analysis and to assess whether the trial actually presents an outlier or agrees with the general findings of the other studies and looking at the weight of the particular study. We see that the answer to all of these questions is, that this study would fall very much into the middle of the overall results that it is not the only study contributing a large number of events and that its influence on the overall estimate of effect is not pulling the effect in one direction or the other.

Who believes the risk of bias is of concern?

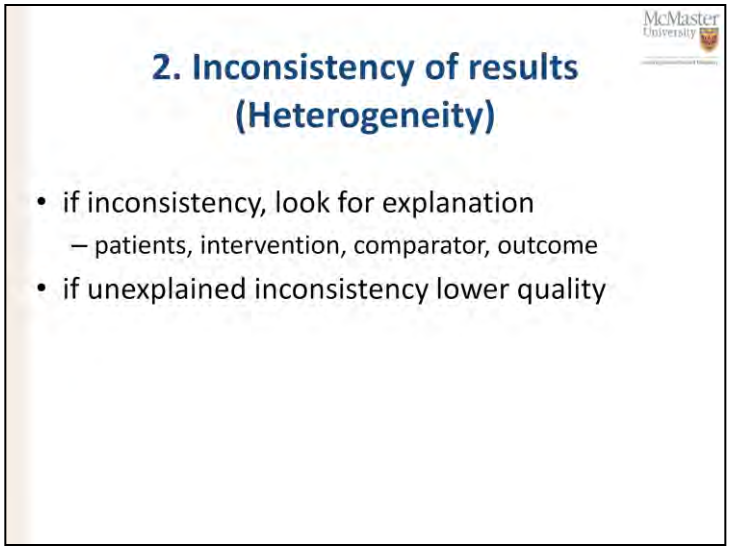
Yes

No

Don’t know or undecided

2. Inconsistency of results (Heterogeneity)

- if inconsistency, look for explanation
  - patients, intervention, comparator, outcome
- if unexplained inconsistency lower quality



Inconsistency of the results or heterogeneity is the second quality criterion. If there is inconsistency one needs to look for an explanation. That is, we can look for whether differences in the population of patients, the intervention, the comparator or the outcome that is how it is measured between studies explain differences in the results across studies. If there is unexplained inconsistency we lower our confidence in the estimate of effect or the quality of the evidence.

McMaster University
Inspiring Innovation and Discovery

Reminders for immunization uptake

Analysis 2.1. Comparison 2 letter reminders vs. control, Outcome 1 Immunized.

Review: Patient reminder and recall systems to improve immunization rates

Comparison: 2 letter reminders vs. control

Outcome: 1 Immunized

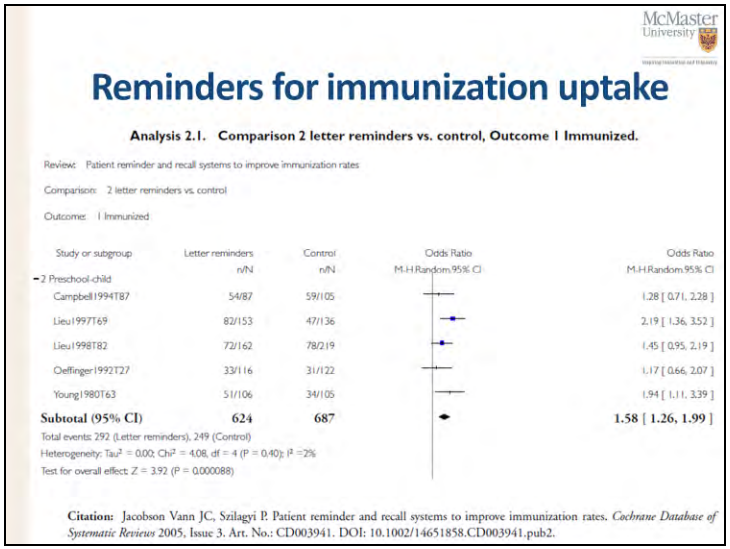
Study or subgroup	Letter reminders n/N	Control n/N	Odds Ratio M-H,Random,95% CI	Odds Ratio M-H,Random,95% CI
2 Preschool-child				
Campbell1994T87	54/87	59/105		1.28 [0.71, 2.28]
Lieu1997T69	82/153	47/136		2.19 [1.36, 3.52]
Lieu1998T82	72/162	78/219		1.45 [0.95, 2.19]
Oeffinger1992T27	33/116	31/122		1.17 [0.66, 2.07]
Young1980T63	51/106	34/105		1.94 [1.11, 3.39]
Subtotal (95% CI)	**624**	**687**		**1.58 [1.26, 1.99]**

Total events: 292 (Letter reminders), 249 (Control)

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 4.08$, $df = 4$ ($P = 0.40$); $I^2 = 2\%$

Test for overall effect: $Z = 3.92$ ($P = 0.000088$)

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.



For example, this first plot is from a body of evidence that looked at whether patient reminders and recalled systems improve immunization rates. The investigators identified 5 studies, all of these studies indicated that reminder systems do increase the uptake of immunization. The confidence intervals of these 5 studies are overlapping. Furthermore, when looking at statistical testing for heterogeneity the paragraph value for heterogeneity is 0.40, making chance a likely explanation for any differences that are observed between studies and the i^2 value ranging from 0 – 100% indicates that true between study variability is unlikely to explain any variability in the results and the variability is likely due to within study variability. While there are no precise thresholds or cut off values for the i^2 guidance indicates that values under of below 50% indicate that heterogeneity is not of great importance. It must be said that these values are not absolute values and they may depend on issues such as sample size.

Analysis 6.1. Comparison 6 patient & provider reminder vs. control, Outcome 1 Immunized.

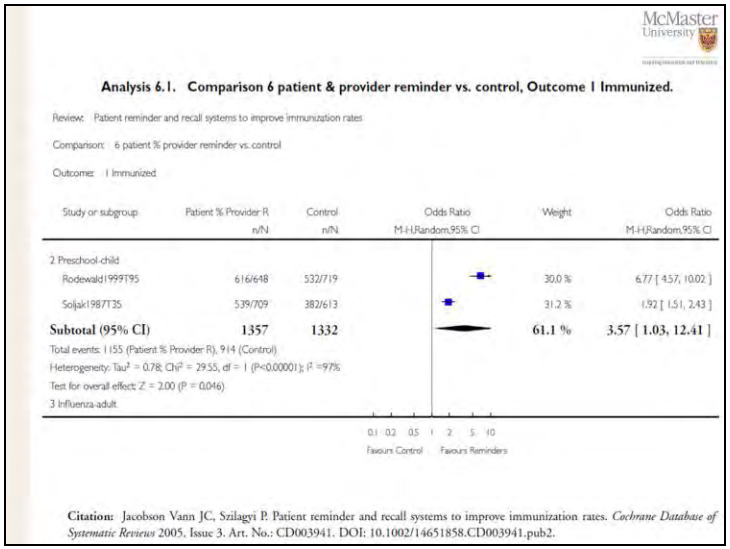
Review: Patient reminder and recall systems to improve immunization rates

Comparison: 6 patient % provider reminder vs. control

Outcome: 1 Immunized

Study or subgroup	Patient % Provider R n/N	Control n/N	Odds Ratio M-H,Random,95% CI	Weight	Odds Ratio M-H,Random,95% CI
2 Preschool-child					
Rodewald1999T95	616/648	532/719		30.0 %	6.77 [4.57, 10.02]
Soljak1987T35	539/709	382/613		31.2 %	1.92 [1.51, 2.43]
Subtotal (95% CI)	**1357**	**1332**		**61.1 %**	**3.57 [1.03, 12.41]**
Total events: 1155 (Patient % Provider R), 914 (Control)					
Heterogeneity: $Tau^2 = 0.78$; $Chi^2 = 29.55$, df = 1 ($P<0.00001$); $I^2 = 97\%$					
Test for overall effect: Z = 2.00 (P = 0.046)					
3 Influenza-adult					

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.



McMaster University

Analysis 6.1. Comparison 6 patient & provider reminder vs. control, Outcome 1 Immunized.

Review: Patient reminder and recall systems to improve immunization rates

Comparison: 6 patient % provider reminder vs. control

Outcome: 1 Immunized

Study or subgroup	Patient % Provider R n/N	Control n/N	Odds Ratio M-H,Random,95% CI	Weight	Odds Ratio M-H,Random,95% CI
2 Preschool-child					
Rodewald1999T95	616/648	532/719		30.0 %	6.77 [4.57, 10.02]
Soljak1987T35	539/709	382/613		31.2 %	1.92 [1.51, 2.43]
Subtotal (95% CI)	**1357**	**1332**		**61.1 %**	**3.57 [1.03, 12.41]**

Total events: 1155 (Patient % Provider R), 914 (Control)

Heterogeneity: $Tau^2 = 0.78$; $Chi^2 = 29.55$, df = 1 (P<0.00001); $I^2 = 97\%$

Test for overall effect: Z = 2.00 (P = 0.046)

3 Influenza-adult

0.1 0.2 0.5 1 2 5 10

Favours Control Favours Reminders

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.

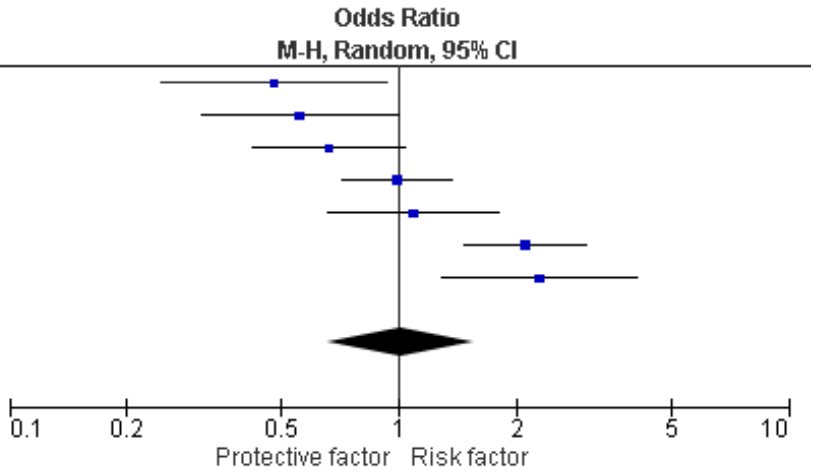
The next slideshows a similar type of intervention. This time two studies were identified for this public health intervention. The two studies show, despite the fact that they both indicate efficacy, widely different results. One study indicates an odds ratio of 6.77, the other and odds ratio of 1.92. While one could say that the intervention is likely to be effective, the actual magnitude of the effect remains uncertain based on the widely differing results here. If for instance our threshold for implementing the intervention was a minimal effect of 3.5 because the intervention comes with significant required resources, we would be left with uncertainty of whether the true effect is really 3.57. And that is based on the fact that the point estimates differ, the confidence intervals are not overlapping, the p value for heterogeneity being very small, and a very large i^2 value. This slide also shows that in the context of decision making heterogeneity is not determined by the fact that the point estimates lie on one side of the relative risk or odds ratio of one.

Non-steroidal drug use and risk of pancreatic cancer

Study or Subgroup	ASA/NSAIDs use Events	Total	No/occasional use Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
Anderson	10	6012	60	17277	12.4%	0.48 [0.24, 0.93]
Menezes	17	79	108	327	13.4%	0.56 [0.31, 1.00]
Ratnasinghe	43	14838	35	7996	14.8%	0.66 [0.42, 1.03]
Jacobs	37	7769	3455	721041	16.1%	0.99 [0.72, 1.38]
Coogan	18	188	207	2339	14.2%	1.09 [0.66, 1.81]
Schernhammer	37	10292	153	89541	15.7%	2.11 [1.47, 3.02]
Langman	25	48	413	1286	13.4%	2.30 [1.29, 4.10]
Total (95% CI)		**39226**		**839807**	**100.0%**	**1.01 [0.65, 1.55]**
Total events	187		4431			

Heterogeneity: $Tau^2 = 0.28$; $Chi^2 = 35.73$, df = 6 ($P < 0.00001$); $I^2 = 83\%$

Test for overall effect: $Z = 0.04$ ($P = 0.97$)



Capurso G, Schünemann HJ, Terrenato I, Moretti A, Koch M, Muti P, Capurso L, Delle Fave G. *Meta-analysis: the use of non-steroidal anti-inflammatory drugs and pancreatic cancer risk for different exposure categories.* Aliment Pharmacol Ther. 2007 Oct 15;26(8):1089-99.

Inconsistency

- I^2
- P-value
- Overlap in CI
- Difference in point estimates

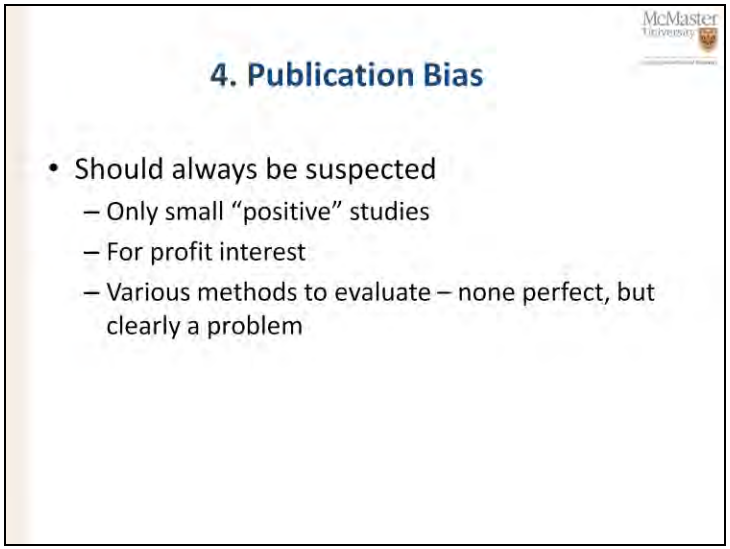
3. Directness of Evidence

generalizability, transferability, applicability

- differences in
  - populations/patients (adults-children)
  - interventions (new vaccine - old)
  - comparator appropriate (placebo – no vaccine – old)
  - outcomes (important – surrogate; immune response – mortality; hepatitis B – liver cancer)
- indirect comparisons
  - interested in A versus B
  - have A versus C and B versus C
  - Rotarix versus no intervention versus RotaTeq versus no intervention

4. Publication Bias

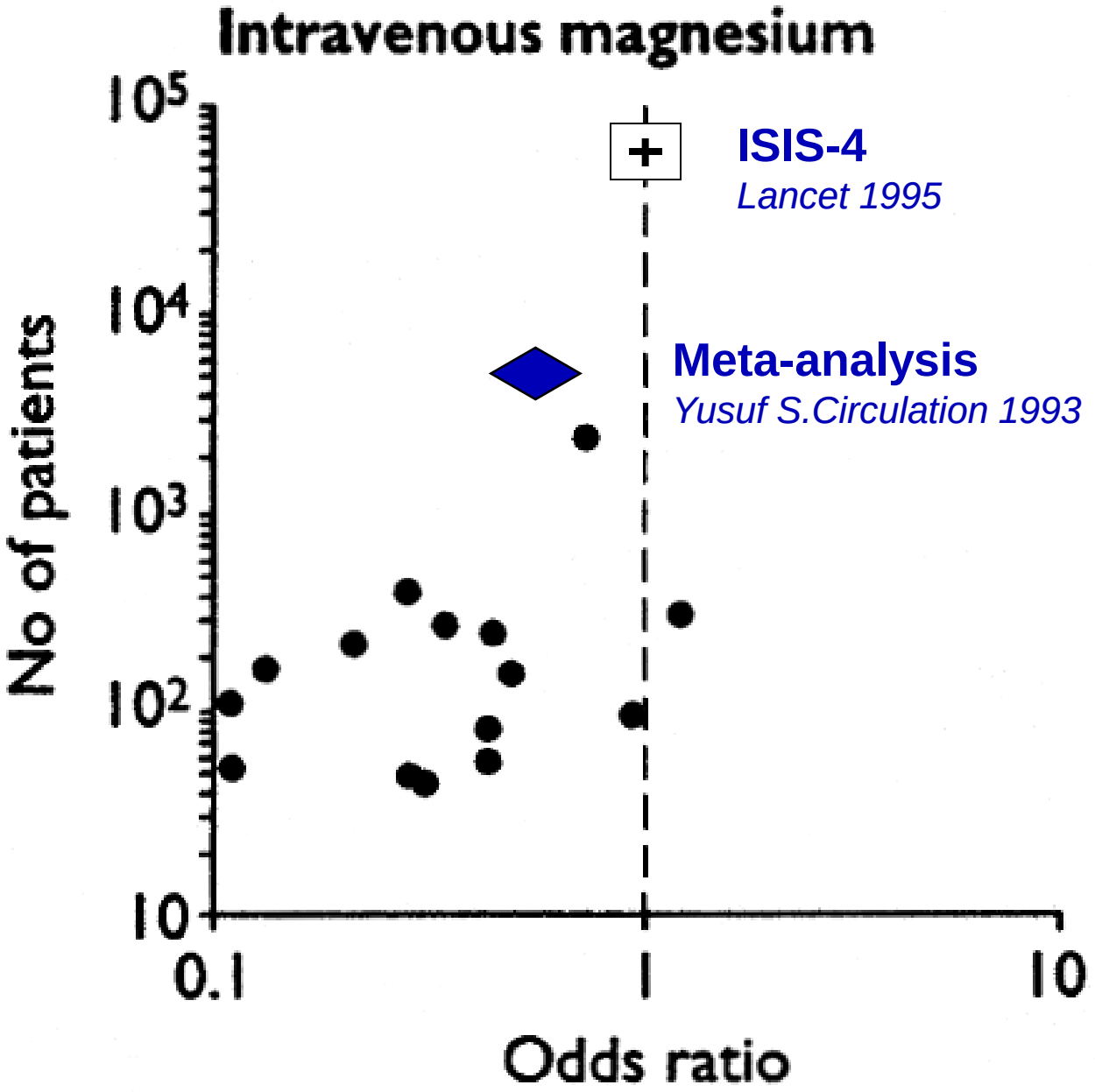
- Should always be suspected
  - Only small “positive” studies
  - For profit interest
  - Various methods to evaluate – none perfect, but clearly a problem



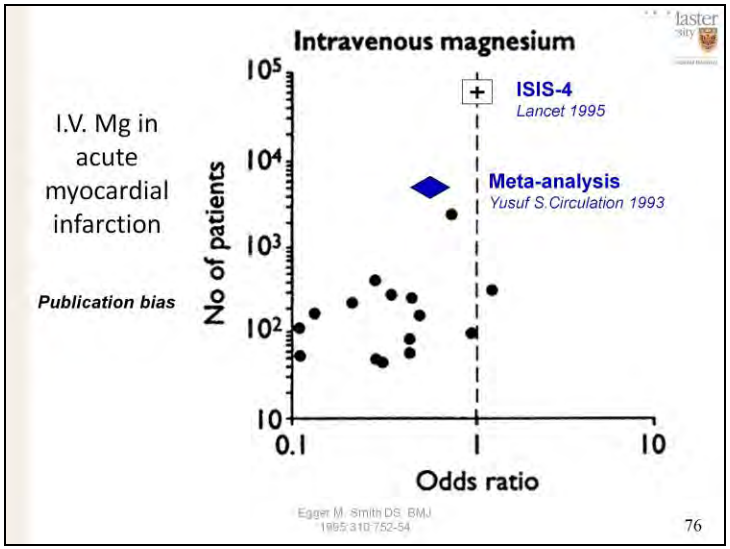
The next factor that may lead to downgrading the confidence and estimates of effect or quality of evidence is publication bias. Publication bias should always be suspected. It refers to the systematic under or over estimate of an effect due to selective publication of studies. It should be suspected in particular when there are only small positive studies, when there is GRADE for profit interest and there are many methods to evaluate publication bias, none of them is perfect but publication bias is clearly a problem. For instance, investigators can use inverted funnel plots to evaluate publication bias.

I.V. Mg in acute myocardial infarction

Publication bias

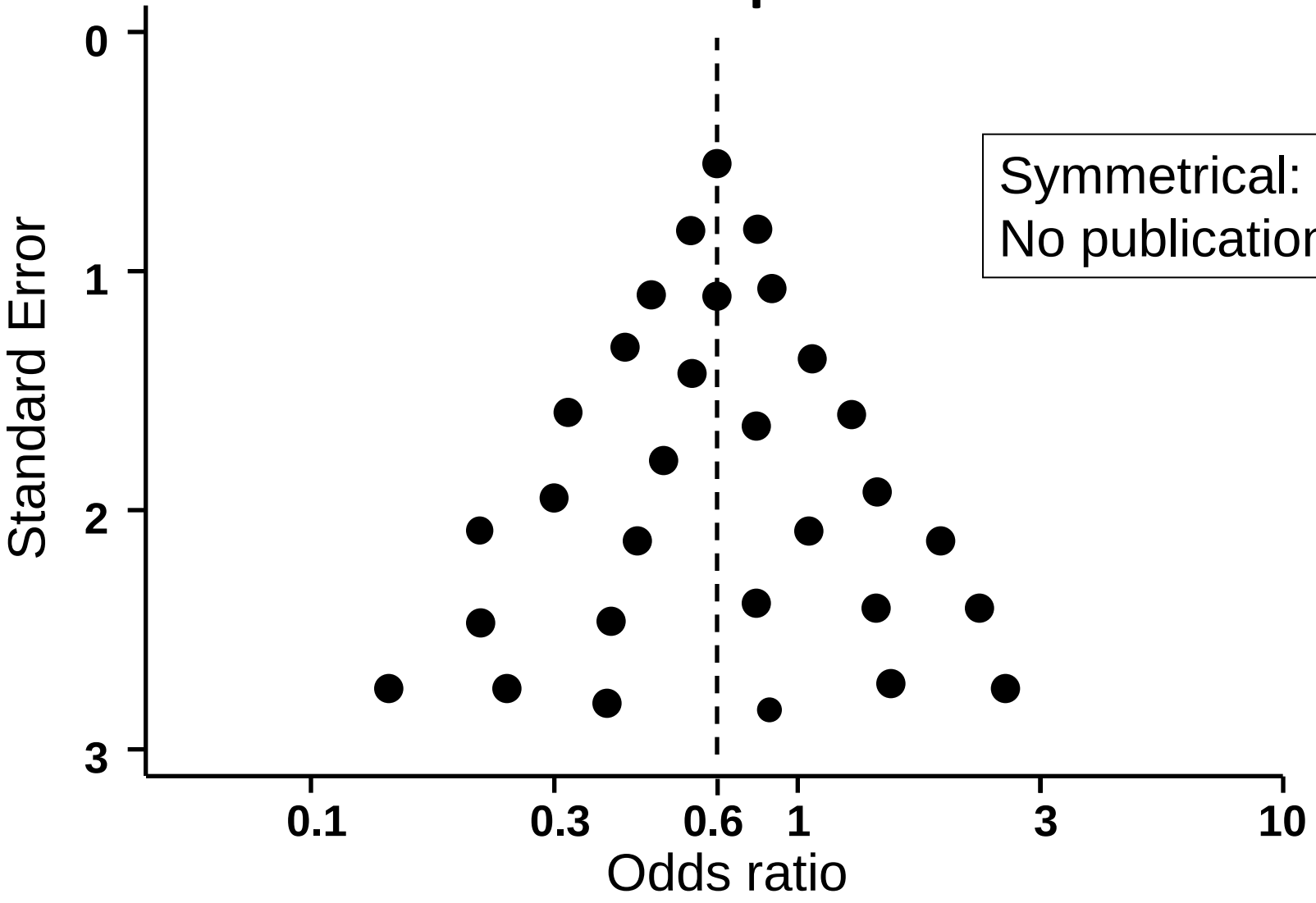


Egger M, Smith DS. BMJ 1995;310:752-54



CHARMAINE CAN YOU PLEASE PULL THIS from prior dictations about publication bias?

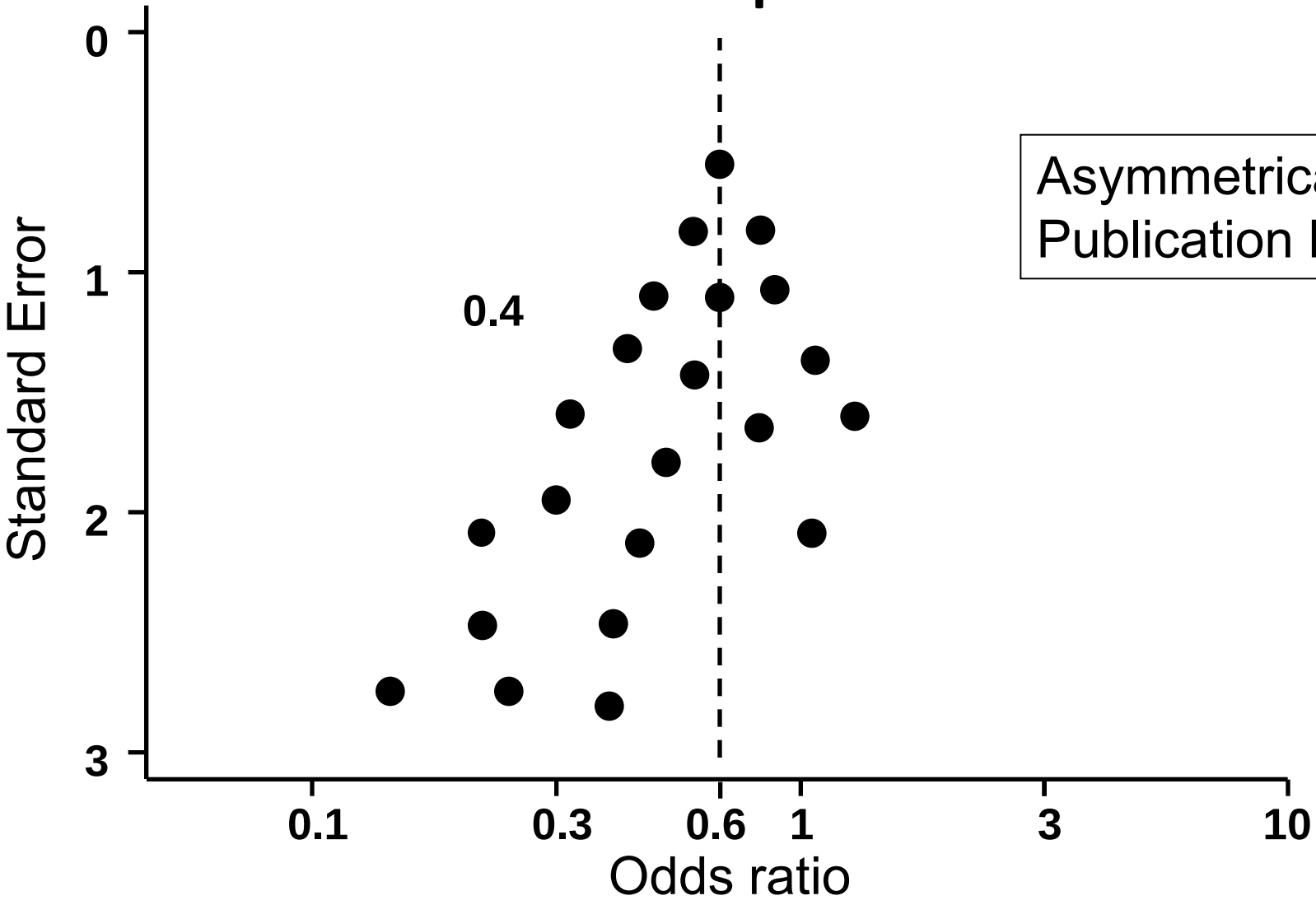
Funnel plot



Symmetrical:
No publication bias

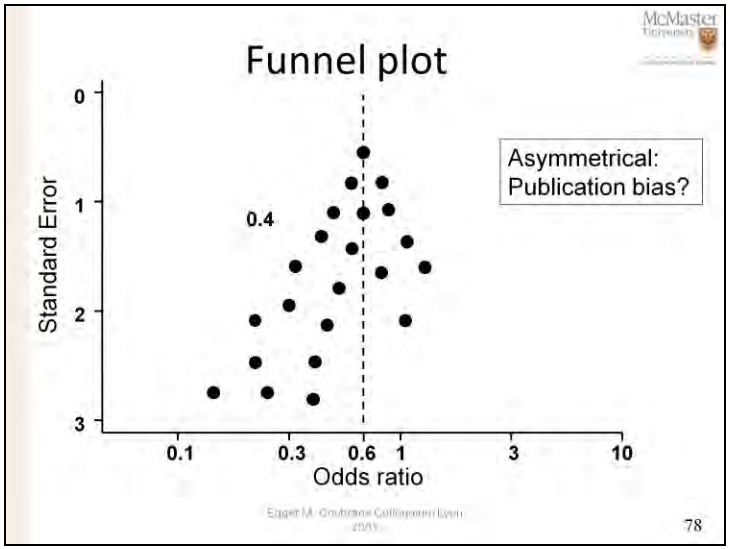
Egger M, Cochrane Colloquium Lyon 2001

Funnel plot



Asymmetrical:
Publication bias?

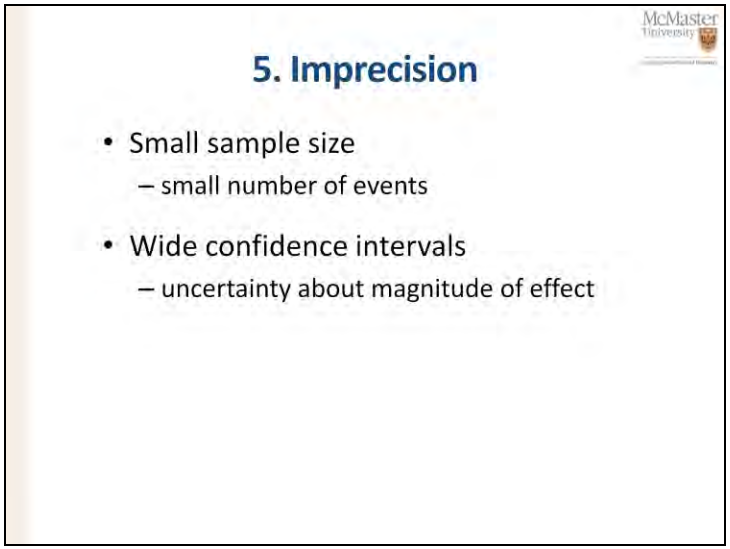
Egger M, Cochrane Colloquium Lyon 2001



The GRADE approach to publication bias is that the quality of evidence for an outcome will be downgraded depending on the degree of publication bias. Publication bias is either labeled as undetected, which does not lead to downgrading, it is strongly suspected, which means downgrading by one level or very strongly suspected, which leads to downgrading by two levels.

5. Imprecision

- Small sample size
  - small number of events
- Wide confidence intervals
  - uncertainty about magnitude of effect



The fifth factor that may lead to the downgrading the quality of evidence is imprecision. It has to do with when there are only very small sample sizes, in particular when there is a small number of events. That usually leads to wide confidence intervals and uncertainty about the magnitude of the true effect.

Example: Immunization in children

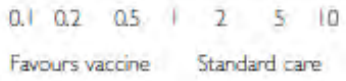
Analysis 4.3. Comparison 4 Inactivated vaccines - (cohort studies by age group), Outcome 3 Otitis media.

Review: Vaccines for preventing influenza in healthy children

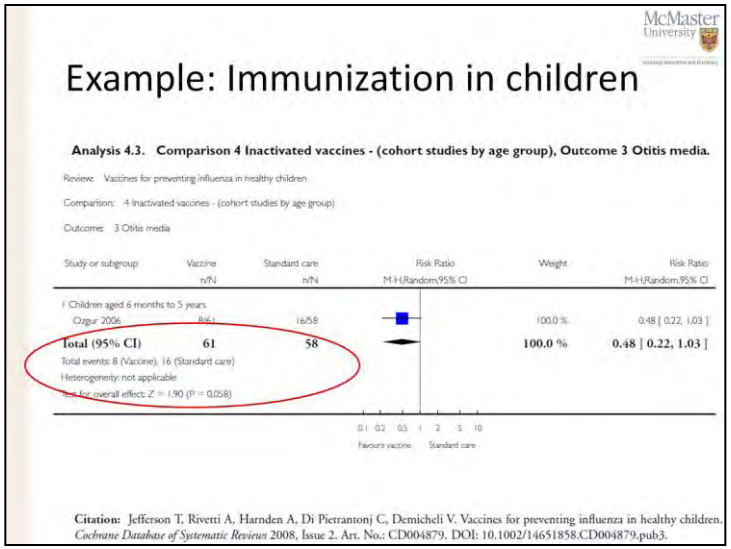
Comparison: 4 Inactivated vaccines - (cohort studies by age group)

Outcome: 3 Otitis media

Study or subgroup	Vaccine n/N	Standard care n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Children aged 6 months to 5 years					
Ozgur 2006	8/61	16/58		100.0 %	0.48 [0.22, 1.03]
Total (95% CI)	**61**	**58**		**100.0 %**	**0.48 [0.22, 1.03]**
Total events: 8 (Vaccine), 16 (Standard care)					
Heterogeneity: not applicable					
Test for overall effect: Z = 1.90 (P = 0.058)					



Citation: Jefferson T, Rivetti A, Harnden A, Di Pietrantonj C, Demicheli V. Vaccines for preventing influenza in healthy children. *Cochrane Database of Systematic Reviews* 2008, Issue 2. Art. No.: CD004879. DOI: 10.1002/14651858.CD004879.pub3.



For example, this first plot shows the inclusion of only one single study that enrolled less than 120 patients and had only 24 events recorded. Despite the large effect, the small number of events and study participants would likely lead to downgrading the quality of evidence by two levels.

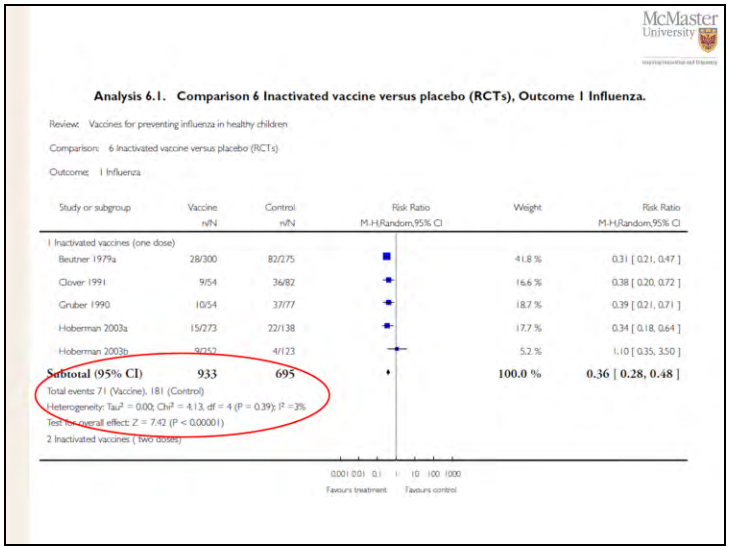
Analysis 6.1. Comparison 6 Inactivated vaccine versus placebo (RCTs), Outcome 1 Influenza.

Review: Vaccines for preventing influenza in healthy children

Comparison: 6 Inactivated vaccine versus placebo (RCTs)

Outcome: 1 Influenza

Study or subgroup	Vaccine n/N	Control n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Inactivated vaccines (one dose)					
Beutner 1979a	28/300	82/275		41.8 %	0.31 [0.21, 0.47]
Clover 1991	9/54	36/82		16.6 %	0.38 [0.20, 0.72]
Gruber 1990	10/54	37/77		18.7 %	0.39 [0.21, 0.71]
Hoberman 2003a	15/273	22/138		17.7 %	0.34 [0.18, 0.64]
Hoberman 2003b	9/252	4/123		5.2 %	1.10 [0.35, 3.50]
Subtotal (95% CI)	**933**	**695**		**100.0 %**	**0.36 [0.28, 0.48]**
Total events: 71 (Vaccine), 181 (Control)					
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 4.13$, df = 4 (P = 0.39); $I^2 = 3\%$					
Test for overall effect: Z = 7.42 (P < 0.00001)					
2 Inactivated vaccines (two doses)					



McMaster University

Analysis 6.1. Comparison 6 Inactivated vaccine versus placebo (RCTs), Outcome 1 Influenza.

Review: Vaccines for preventing influenza in healthy children

Comparison: 6 Inactivated vaccine versus placebo (RCTs)

Outcome: 1 Influenza

Study or subgroup	Vaccine n/N	Control n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Inactivated vaccines (one dose)					
Beutner 1979a	28/300	82/275		41.8 %	0.31 [0.21, 0.47]
Clover 1991	9/54	36/82		16.6 %	0.38 [0.20, 0.72]
Gruber 1990	10/54	37/77		18.7 %	0.39 [0.21, 0.71]
Hoberman 2003a	15/273	22/138		17.7 %	0.34 [0.18, 0.64]
Hoberman 2003b	9/252	4/123		5.2 %	1.10 [0.35, 3.50]
Subtotal (95% CI)	**933**	**695**		**100.0 %**	**0.36 [0.28, 0.48]**
Total events: 71 (Vaccine), 181 (Control)					
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 4.13$, df = 4 (P = 0.39); $I^2 = 3\%$					
Test for overall effect: Z = 7.42 (P < 0.00001)					
2 Inactivated vaccines (two doses)					

0.001 0.01 0.1 1 10 100 1000

Favours treatment Favours control

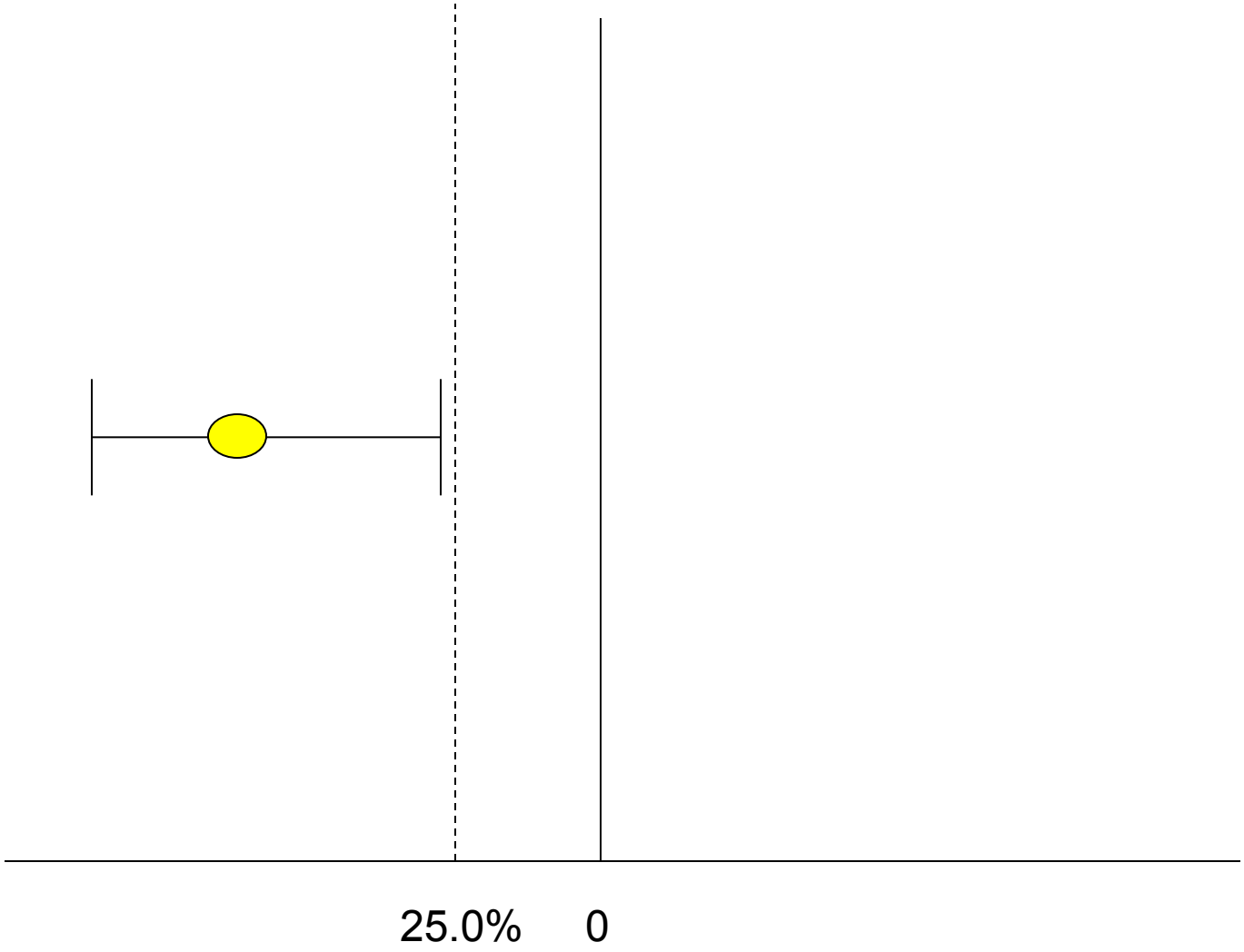
The next example shows a systematic review that included five studies. Of note, one of the studies is not statistically significant however in GRADE we look at impression across studies such as we do for the other factors that lead to downgrading the quality of evidence or upgrading the quality of evidence. An imprecise single study would not influence the judgment. We would look at the overall results and quickly realize that there were approximately 1700 individuals enrolled in these studies, there were about 250 events, 252 to be exact, and the confidence intervals around the point estimate of 0.36 for the risk ratio is very tight. Evidence such as that would not be downgraded for imprecision, given the large number of events, the tight confidence interval and the relatively large sample size.

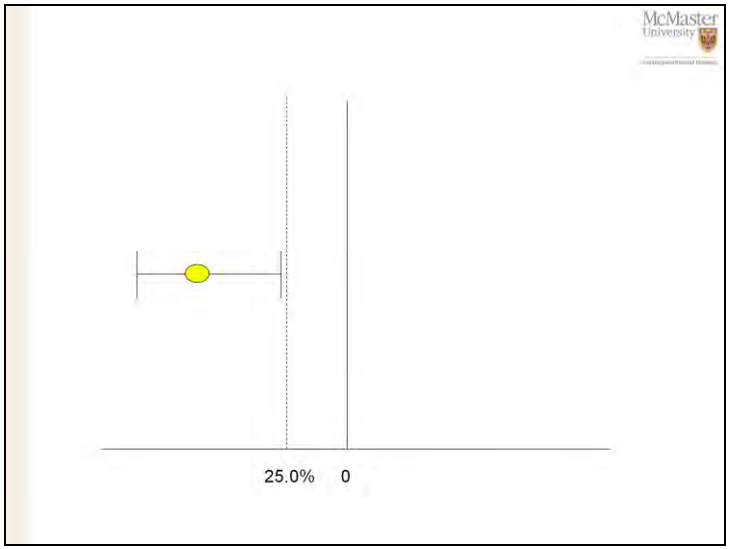
For systematic reviews

- If the 95% CI excludes a relative risk (RR) of 1.0 and the total number of events or patients exceeds the OIS criterion, precision is adequate. If the 95% CI includes appreciable benefit or harm (we suggest a RR of under 0.75 or over 1.25 as a rough guide) rating down for imprecision may be appropriate even if OIS criteria are met.

Optimal information size

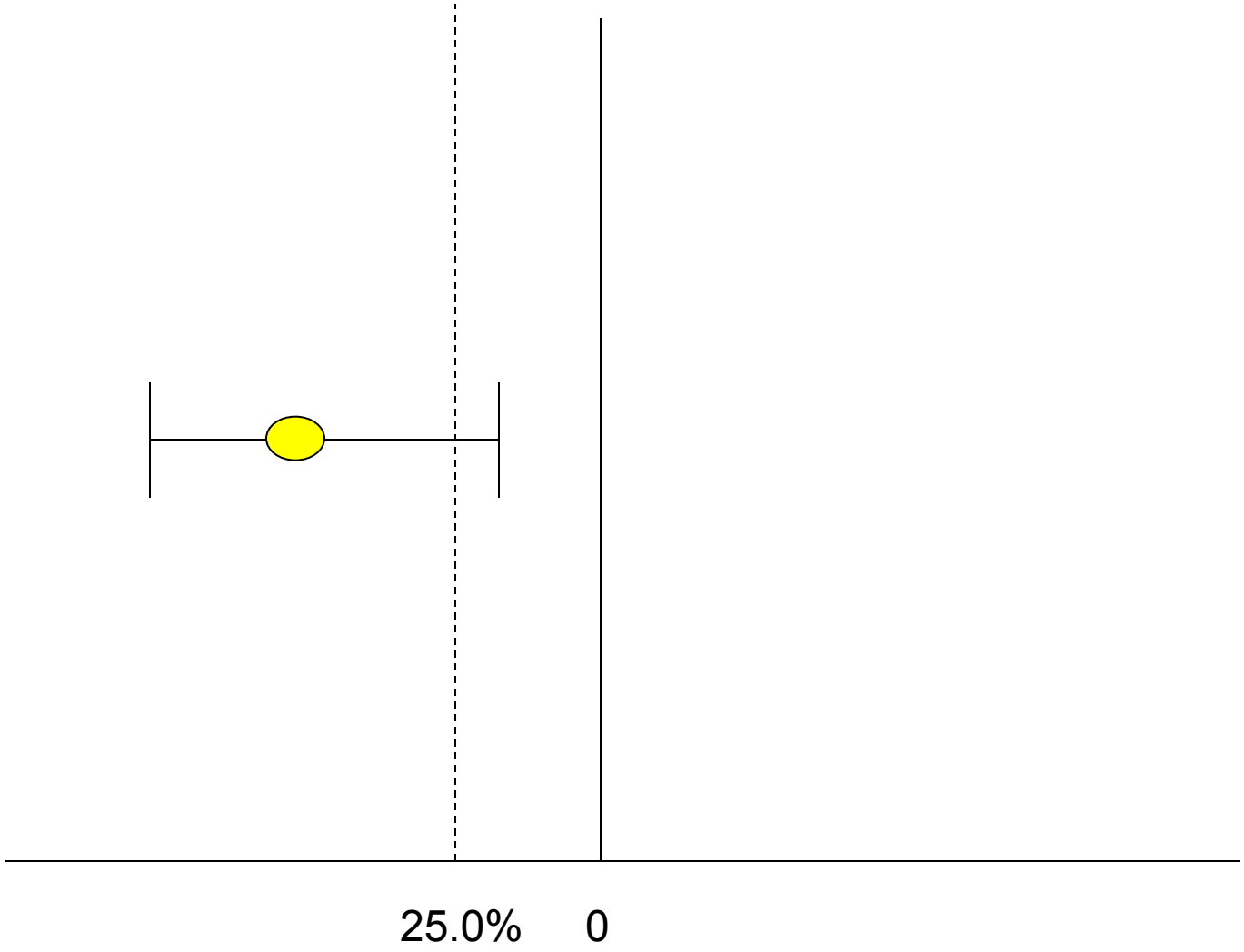
- We suggest the following: if the total number of patients included in a systematic review is less than the number of patients generated by a conventional sample size calculation for a single adequately powered trial, consider rating down for imprecision. Authors have referred to this threshold as the “optimal information size” (OIS)

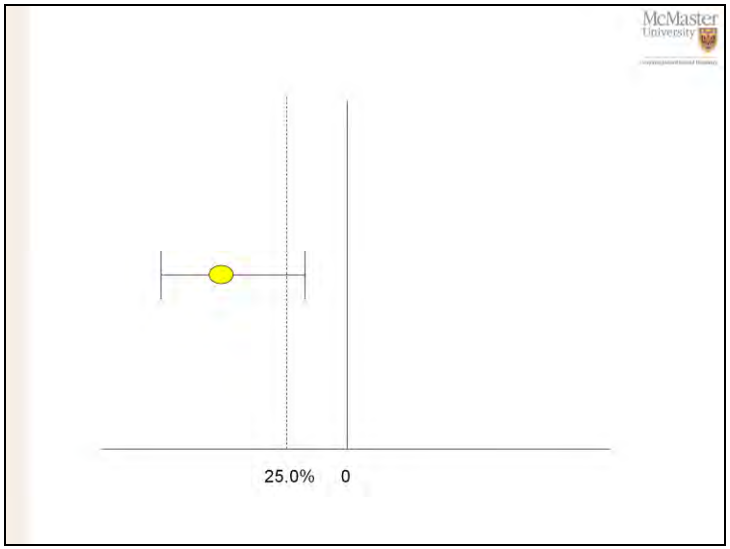




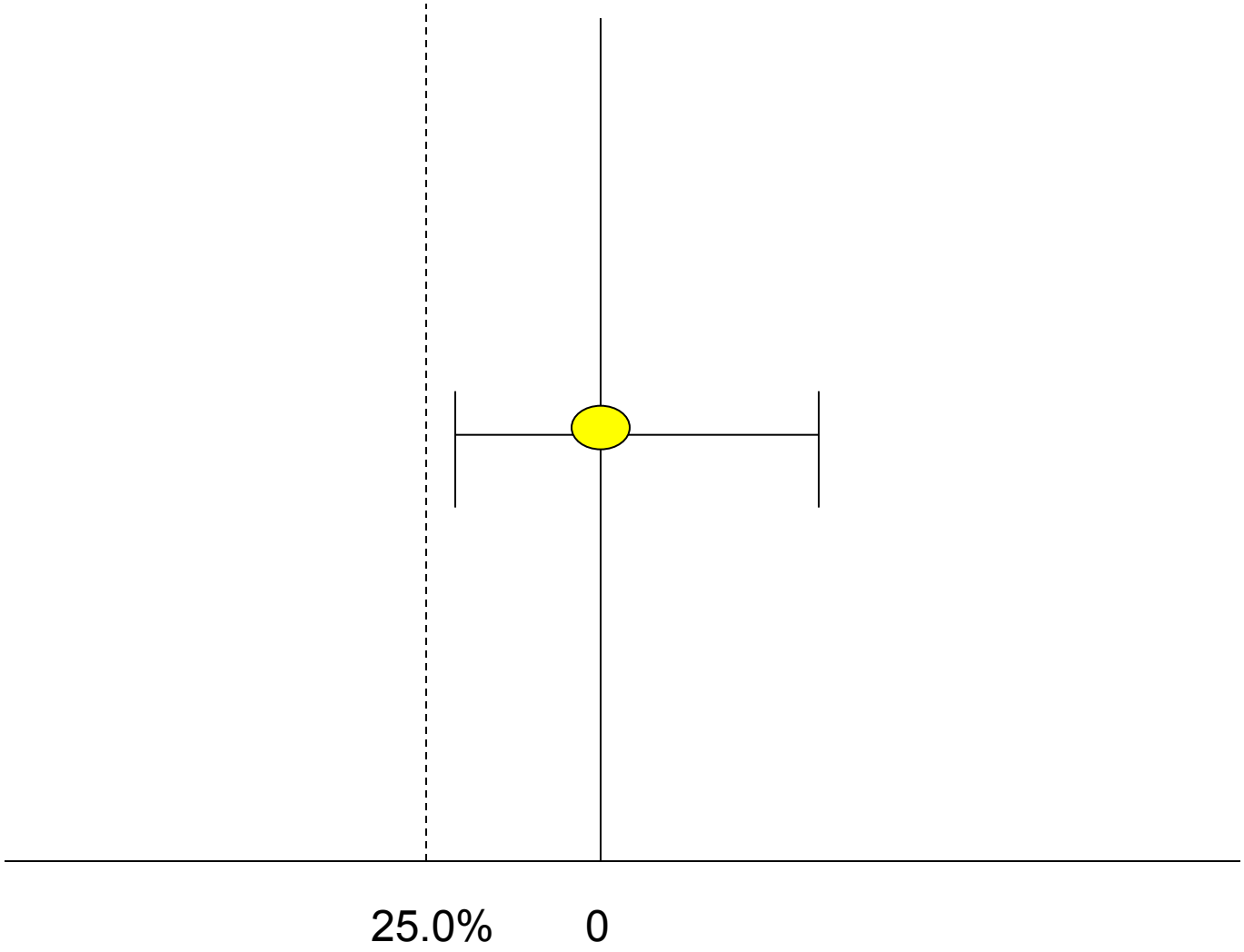
This can be made clearer, so under -- If you were to consider this a relative risk reduction of 25%, this would be a relative risk reduction of 0% -- so, no effect.

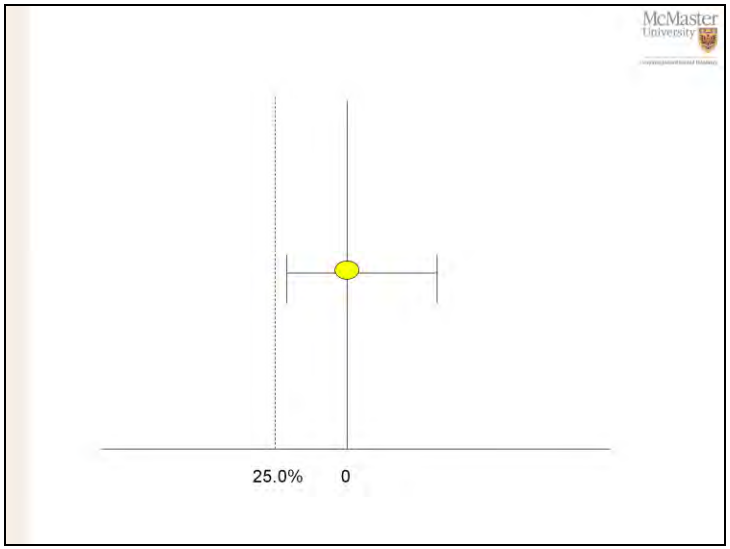
If you find a result that looks like this in your meta-analysis, to a point estimate that is larger than a 25% risk reduction, confidence intervals not overlapping, it's pretty clear-cut -- the results are not imprecise.



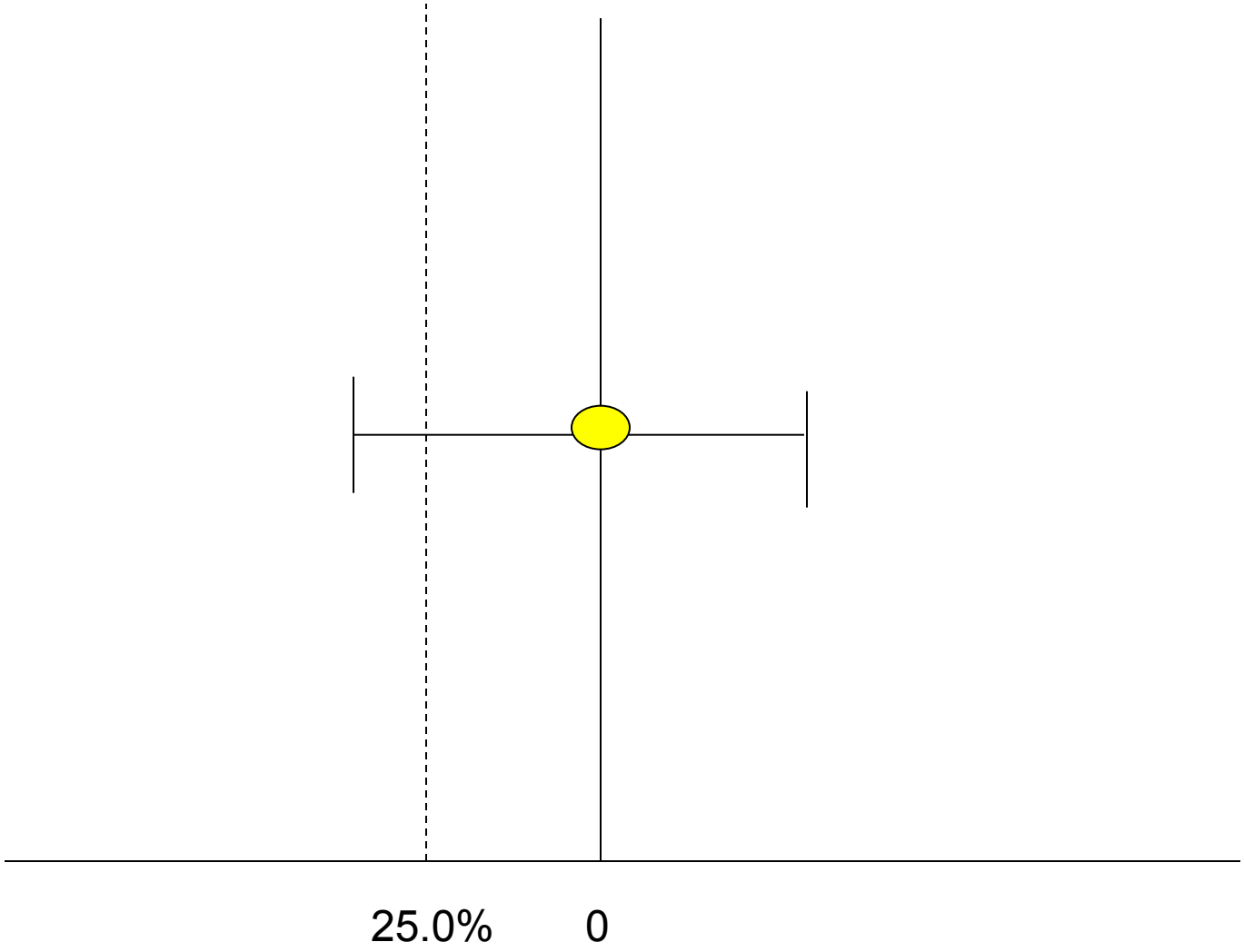


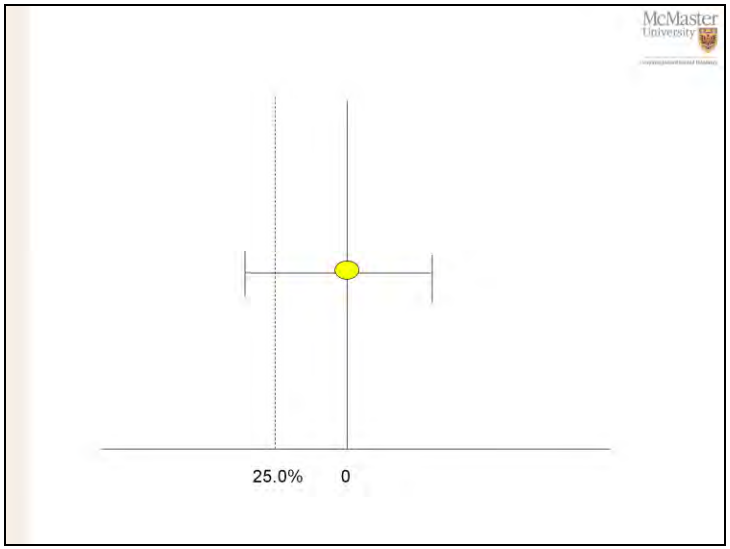
If you find something like that and your threshold for relative risk reduction is really 25% -- and this is what you would need to achieve in order to be confident that the results are precise enough -- despite the fact that they may be statistically significant, you may rate down for imprecision because you really are not confident that the effect that you would try to achieve is achieved.





At the same time -- this is the example that I described -- you may see no effect of an intervention, and the confidence interval may be relatively narrow and not include what we use as a rough guide -- the 25% relative risk reduction. You may say, "This is precise enough.





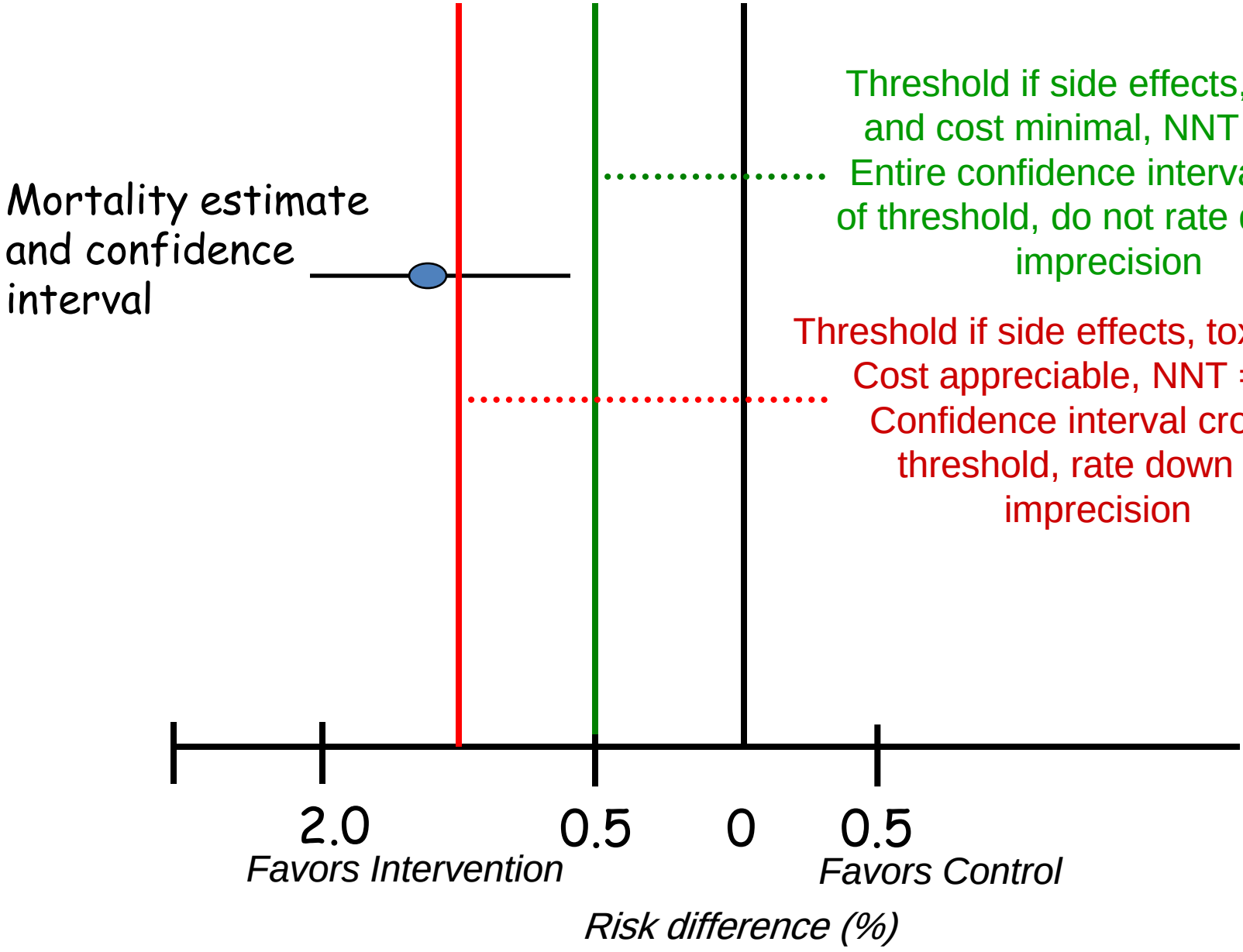
"We don't expect additional information to change this dramatically," as opposed to a situation like this, where, despite the fact that you have no effect, your confidence interval still includes the possibility of an appreciable benefit or harm.

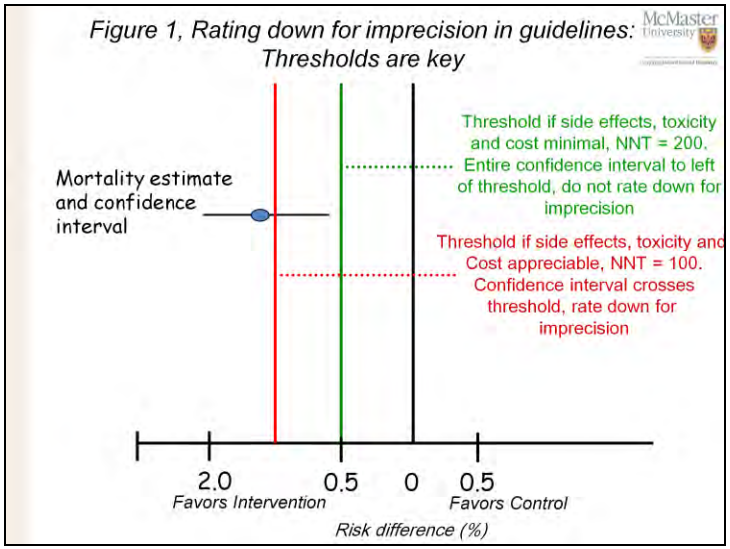
Under those circumstances, you really are not very confident that you can really say that there is no effect.

And the issue, then, when you go to guidelines, becomes that your thresholds are becoming key.

Figure 1, Rating down for imprecision in guidelines: Thresholds are key

McMaster University
Inspiring Innovation and Discovery





The thresholds usually are based on absolute estimates of effect.

So, just to take you through this relatively quickly -- So, if, for instance, you would see mortality estimates as follows -- so, these are absolute estimates of effect, the risk difference of 2%, .05%, 0%, and a 0.5% increase.

So, let's assume that your threshold for applying an intervention would be a risk difference of 0.5% -- so, 0.5% or one out of 200 people who would receive the intervention -- die less.

And if your true estimate of effect was the following -- right?

-- so, this is including thresholds -- was the following, you would say, "Okay, I have enough information.

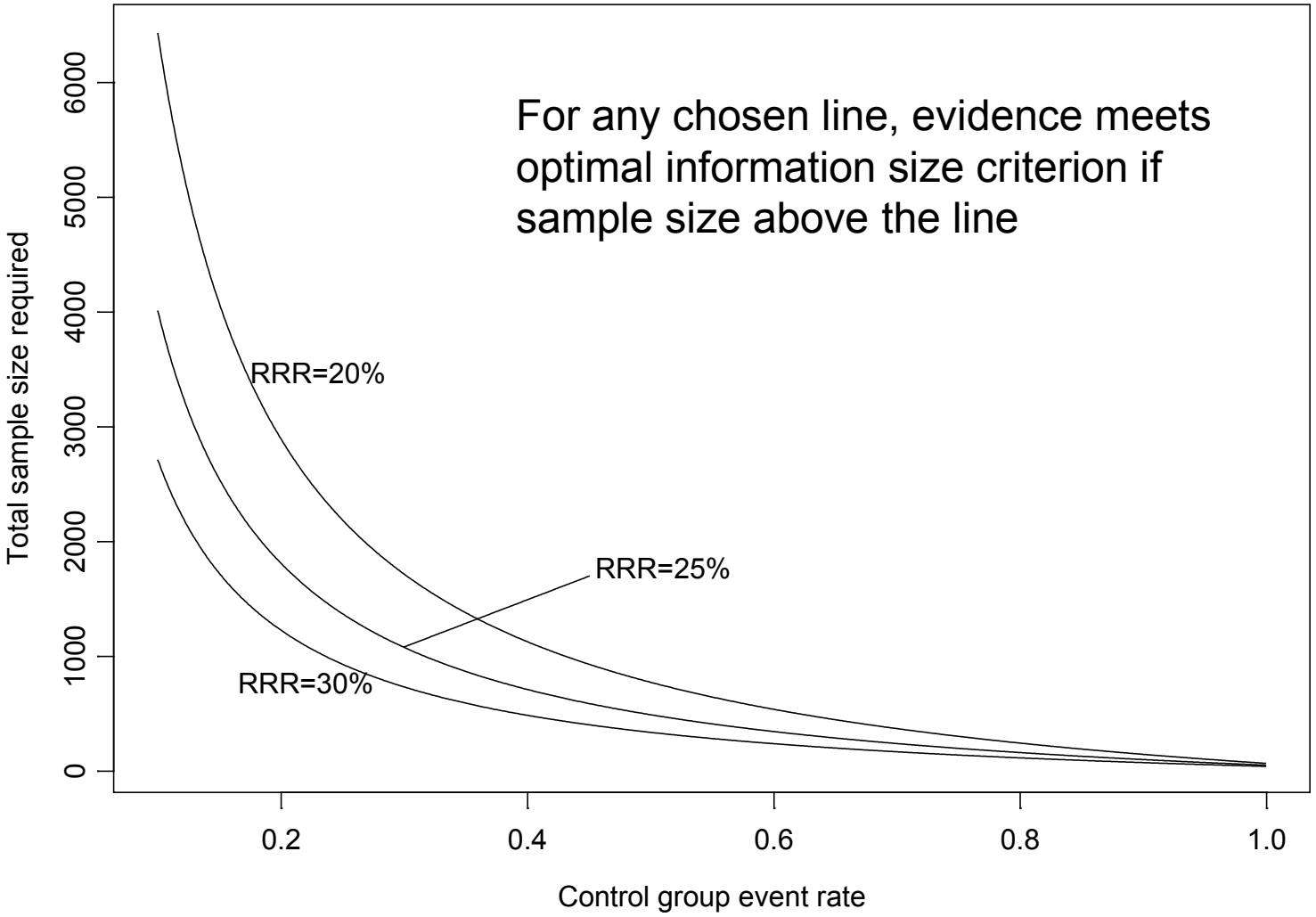
"I'm pretty confident that these estimates of effect "are good enough for me to say that we don't need to

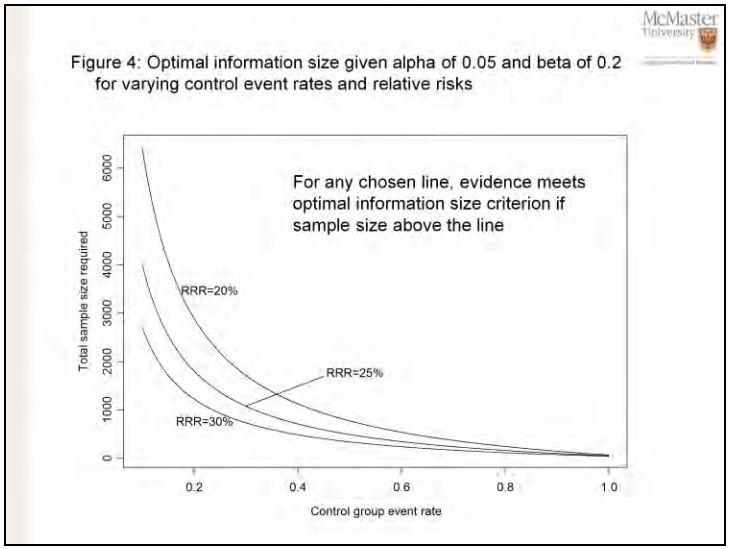
downgrade." If your threshold, however, because of cost, downsides, and other side effects, would be a risk reduction of approximately 1.25% -- 1%, sorry -- which comes with an NNT of 100 -- sorry -- yes, 100 -- excuse me.

So, a risk difference of 1%, and if your true estimate of effect was the following, despite it showing benefit, it would cross this line.

You may still seek more information, or you would ask for more information, and you might downgrade for the quality of evidence.

Figure 4: Optimal information size given alpha of 0.05 and beta of 0.2 for varying control event rates and relative risks





These are curves that we've produced which basically tell you about the optimal information size, and you can see where your body of evidence actually falls on these curves.

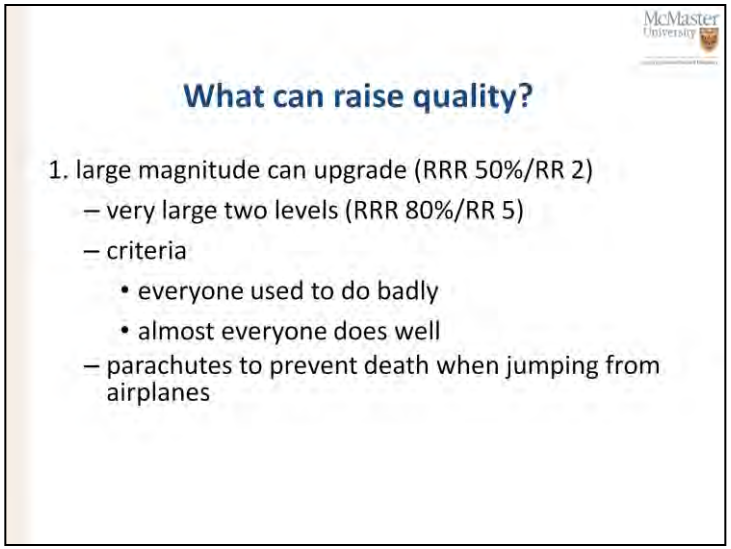
What it explains is, if you are above the line, the optimal information size criteria are met for the various relative estimates of effect, the control group event rate, and the total sample size.

This is fairly easy to apply if you use this as a rough guide.

This will hopefully help with making judgments about precision and imprecision

What can raise quality?

1. large magnitude can upgrade (RRR 50%/RR 2)
   - very large two levels (RRR 80%/RR 5)
   - criteria
     - everyone used to do badly
     - almost everyone does well
   - parachutes to prevent death when jumping from airplanes



There are three factors that can lead to upgrading the quality of evidence. The first is a very large, or large magnitude of effect. We typically use a relative risk reduction of 50% or relative risk of 2, as a threshold of upgrading by one level and the relative risk reduction of 80% or relative risk of 5 as a threshold of upgrading by two levels. It is clear that there may be absolute effects, rather than relative effects that may make us certain that a large effect exists, but we have not defined thresholds for that. One can look at this under the following category that is if there is an intervention, after which almost everyone who would usually do badly, now does well. The example that was mentioned earlier about parachutes to prevent death when jumping from an airplane is a good example for that.

McMaster University
Inspiring Innovation and Discovery

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

HULTON/GETTY

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

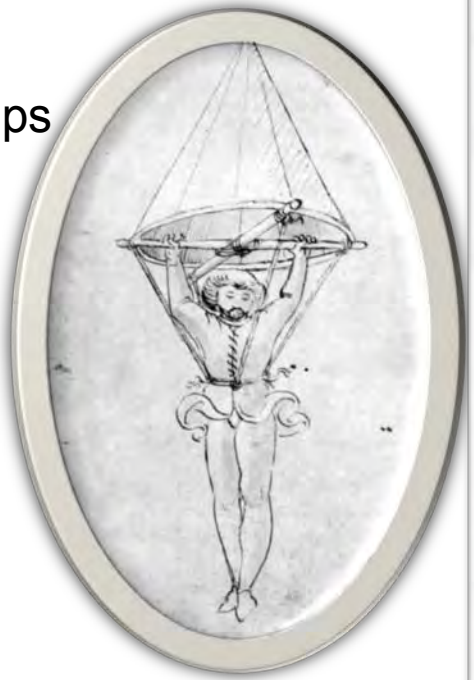
BMJ, 2003

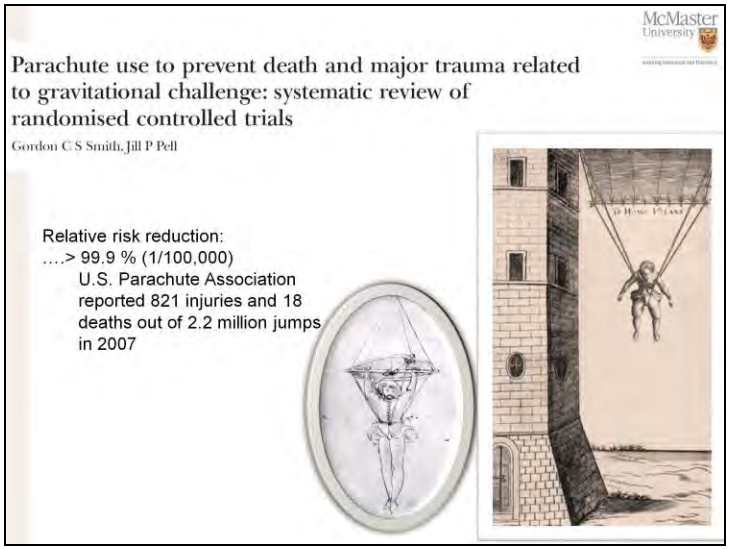
McMaster University
Inspiring Innovation and Discovery

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Relative risk reduction:
….> 99.9 % (1/100,000)
U.S. Parachute Association reported 821 injuries and 18 deaths out of 2.2 million jumps in 2007





The example that was mentioned earlier about parachutes to prevent death when jumping from an airplane is a good example for that.

Reminders for immunization uptake

Review: Patient reminder and recall systems to improve immunization rates

Comparison: 7 Patient Reminders (summary) vs. control

Outcome: 1 Immunized

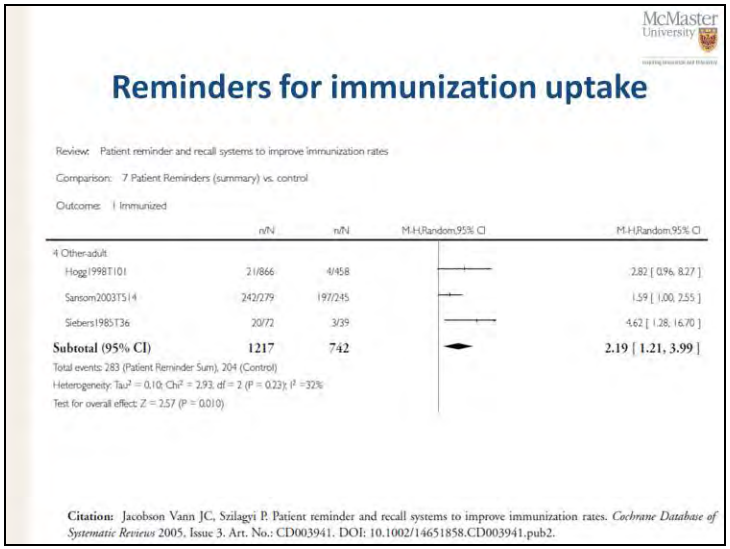
	n/N	n/N	M-H,Random,95% CI	M-H,Random,95% CI
4 Other-adult				
Hogg1998T101	21/866	4/458		2.82 [0.96, 8.27]
Sansom2003T514	242/279	197/245		1.59 [1.00, 2.55]
Siebers1985T36	20/72	3/39		4.62 [1.28, 16.70]
Subtotal (95% CI)	**1217**	**742**		**2.19 [1.21, 3.99]**

Total events: 283 (Patient Reminder Sum), 204 (Control)

Heterogeneity: $Tau^2 = 0.10$; $Chi^2 = 2.93$, df = 2 (P = 0.23); $I^2 = 32\%$

Test for overall effect: Z = 2.57 (P = 0.010)

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.

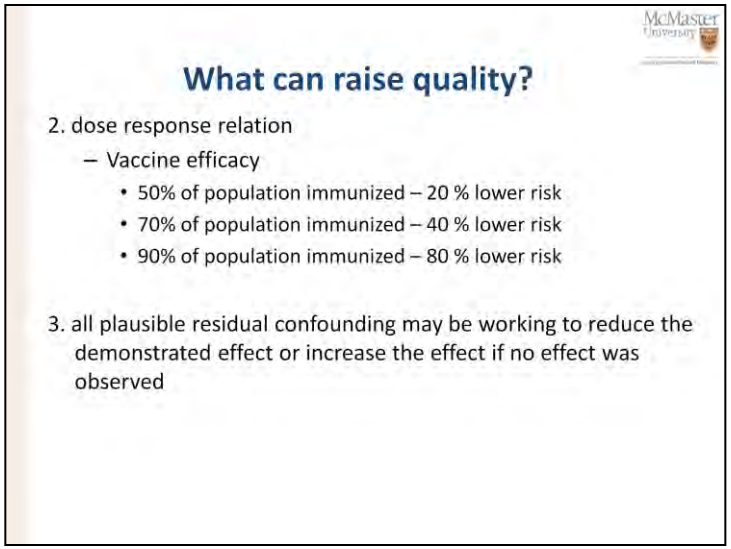


Another example is shown on this slide where the intervention was the provision of patient reminders. There were three studies included in this systematic review, the overall estimate of effect is a relative risk of 2.19 with confidence intervals that probably will fulfill our rules for imprecision where there are approximately 487 events in three studies that enrolled nearly 2000 patients. An effect such as here of 2.19 with these relatively narrow confidence intervals would likely lead us to upgrade the quality of evidence from observational studies by one level. Note that the factor for upgrading the quality of evidence, usually apply to observational studies only.

What can raise quality?

2. dose response relation
   - Vaccine efficacy
     - 50% of population immunized – 20 % lower risk
     - 70% of population immunized – 40 % lower risk
     - 90% of population immunized – 80 % lower risk

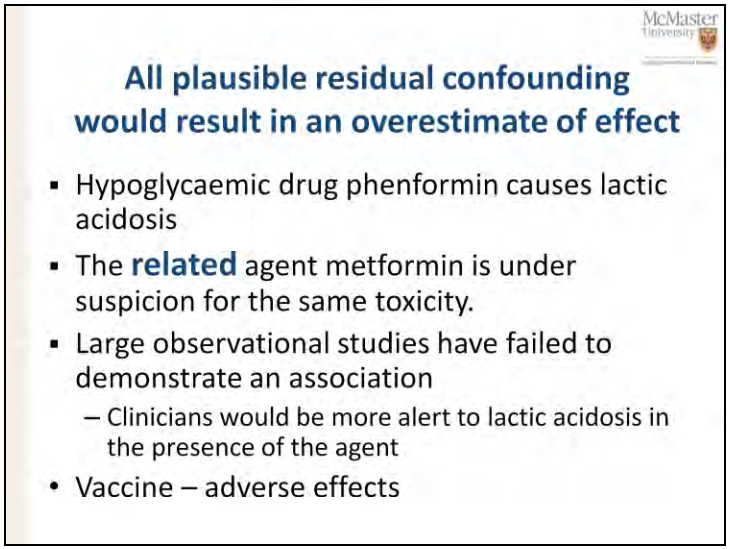
3. all plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed



A second factor that can raise the quality of evidence relates to dose response relations. A hypothetical example comes from observations of population base dose response relations in the context of XXXX efficacy. Imagine a 20% lower risk if 50% of the population is immunized a 40% lower risk of a disease if 70% of the population is immunized and an 80% lower risk if 90% of the population is immunized. Such a dose response relations would make us more confidence that efficacy of the vaccine truly exists; in particular if such an observation is available across different settings and populations. The third factor that can lead to upgrading the quality of evidence relates to if all plausible residual confounding or biases may be working to reduce the demonstrated effect or increase an effect if no effect was observed. The next slide will demonstrate that based on an example.

All plausible residual confounding would result in an overestimate of effect

- Hypoglycaemic drug phenformin causes lactic acidosis
- The **related** agent metformin is under suspicion for the same toxicity.
- Large observational studies have failed to demonstrate an association
  - Clinicians would be more alert to lactic acidosis in the presence of the agent
- Vaccine – adverse effects



Take the situation of the MMR vaccine and the suspected association with autism. If we imagine that there was an earlier report that connected autism to MMR vaccination, it is very likely that subsequently there was a large degree of over reporting of autism after a vaccine had been administered. Despite this over reporting, that is despite the opposing plausible bias and confounding, no association was observed when reviews were done that looked at large observational studies evaluating this association. Under those circumstances, we may confidently increase the quality of the evidence that there truly is no association and this is confirmed by the withdrawal of the early publication that led to this suspected association.

Table 1 Bradford Hill criteria of causality and their relation to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria for upgrading and downgrading

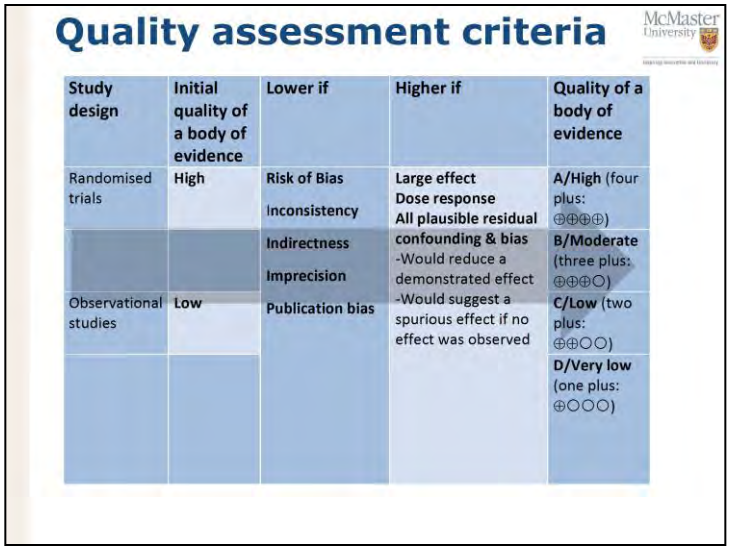
Bradford Hill criteria	Consideration in GRADE
Strength	Strength of association and imprecision in effect estimate
Consistency	Consistency across studies, ie, across different situations (different researchers)
Temporality	Study design, specific study limitations; RCTs fulfil this criterion better than observational studies, properly designed and conducted observational studies
Biological gradient	Dose–response gradient
Specificity	Indirectness
Biological plausibility	Indirectness
Coherence	Indirectness
Experiment	Study design, randomisation, properly designed and conducted observational studies
Analogy	Existing association for critical outcomes will lead to not downgrading the quality, indirectness

Schünemann et al. JECH 2010

Quality assessment criteria

McMaster University
Inspiring Innovation and Discovery

Study design	Initial quality of a body of evidence	Lower if	Higher if	Quality of a body of evidence
Randomised trials	**High**	**Risk of Bias** **Inconsistency**	**Large effect** **Dose response** **All plausible residual confounding & bias** -Would reduce a demonstrated effect -Would suggest a spurious effect if no effect was observed	**A/High** (four plus: ⊕⊕⊕⊕)
		Indirectness **Imprecision**		**B/Moderate** (three plus: ⊕⊕⊕◯)
Observational studies	**Low**	**Publication bias**		**C/Low** (two plus: ⊕⊕◯◯)
				D/Very low (one plus: ⊕◯◯◯)

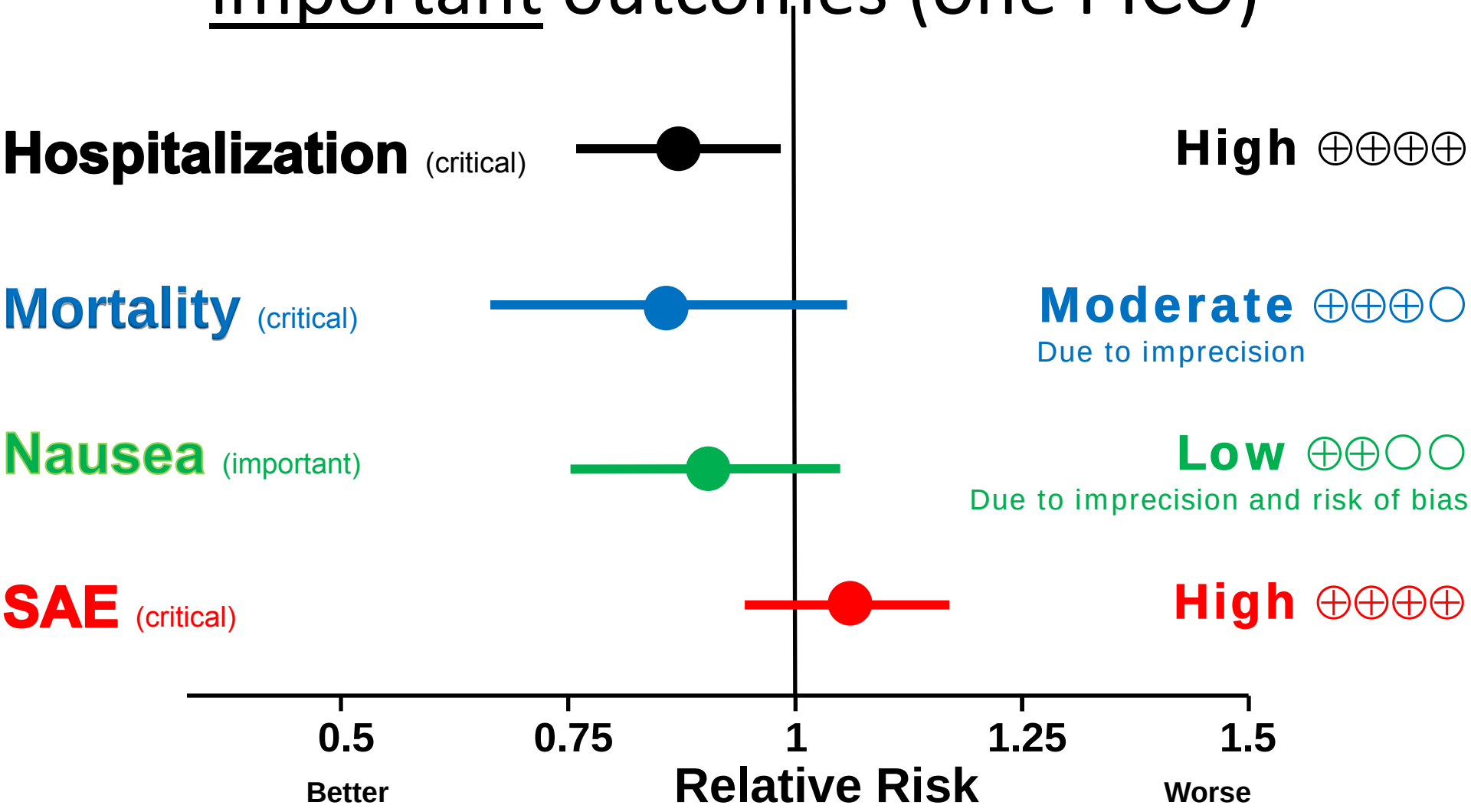


So, in summary, the quality of the evidence or the confidence in an estimate of effect is assessed according to the following criteria. A body of evidence from randomised trials starts as high quality, a body of evidence from observational studies starts as low quality, however there are five factors that in particular for randomized control trials lead to lowering the quality of evidence; those are the risk of bias, inconsistency, indirectness, impression and publication bias. The quality of evidence may be increased if one of the three factors that are listed here is present, a large effect dose response relation or if all plausible residual confounding and biases would oppose the observed effect. The quality of a body of evidence for an outcome is then categorized into one of four categories going from high or 4+ to very low or 1+.

Overall quality of a body of evidence

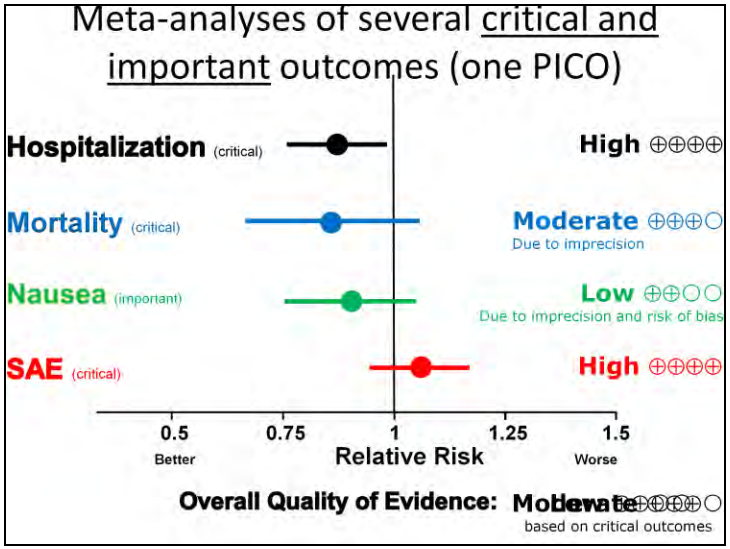
- The quality of evidence reflects the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.
- Guideline developers must specify and determine importance of all relevant outcomes
- Overall quality of evidence is based on the lowest quality of all critical outcomes

Meta-analyses of several critical and important outcomes (one PICO)



Overall Quality of Evidence: **Moderate** ⊕⊕⊕◯

based on critical outcomes



The overall quality of the evidence reflects the extent then of our confidence that the estimates of an effect, as I said, are adequate to support a particular decision or recommendation.

Guideline developers must specify and determine the importance of all relevant outcomes in our view.

And the overall quality, as I also said earlier, of evidence is based on the lowest quality of all critical outcomes.

Now, let's assume this frequently comes up, and it has to do with concerns that we over-penalize or that we are too severe, too stringent in the application of these criteria.

So, let's assume that we have a systematic review, a meta-analysis of several critical and important outcomes.

Okay?

And the fourth outcome, that is serious adverse events.

It's critical, and it's considered high because the confidence interval is considered to be narrow enough under those circumstances.

So based on what I said before, what would be the overall quality of the evidence?

Moderate.

Why is it moderate?

It's the lowest critical, right?

It's quite straightforward.

It's the lowest critical, so, yes.

So, the overall quality of evidence is not low because nausea, despite the fact that it is only low quality, it was rated as important and not critical.

So moderate.

The intervention may just be any intervention.

And hospitalizations were considered to be a critical outcome.

And this is what you would find.

You would find a risk reduction for hospitalizations.

No downgrading takes place.

It's high-quality evidence for hospitalizations.

Let's assume that you have a second outcome, which is mortality.

It is considered critical.

And the quality here is moderate, and perhaps this is due to imprecision because you're not entirely sure whether immortality's increased or decreased over the other effects of mortality.

And let's now also assume that you have a third outcome that is rated as important, but not critical, which is nausea.

And it comes with the following estimate of effect.

And the fourth outcome, that is serious adverse events.

It's critical, and it's considered high because the confidence interval is considered to be narrow enough under those circumstances.

So based on what I said before, what would be the overall quality of the evidence?

Moderate.

Why is it moderate?

It's the lowest critical, right?

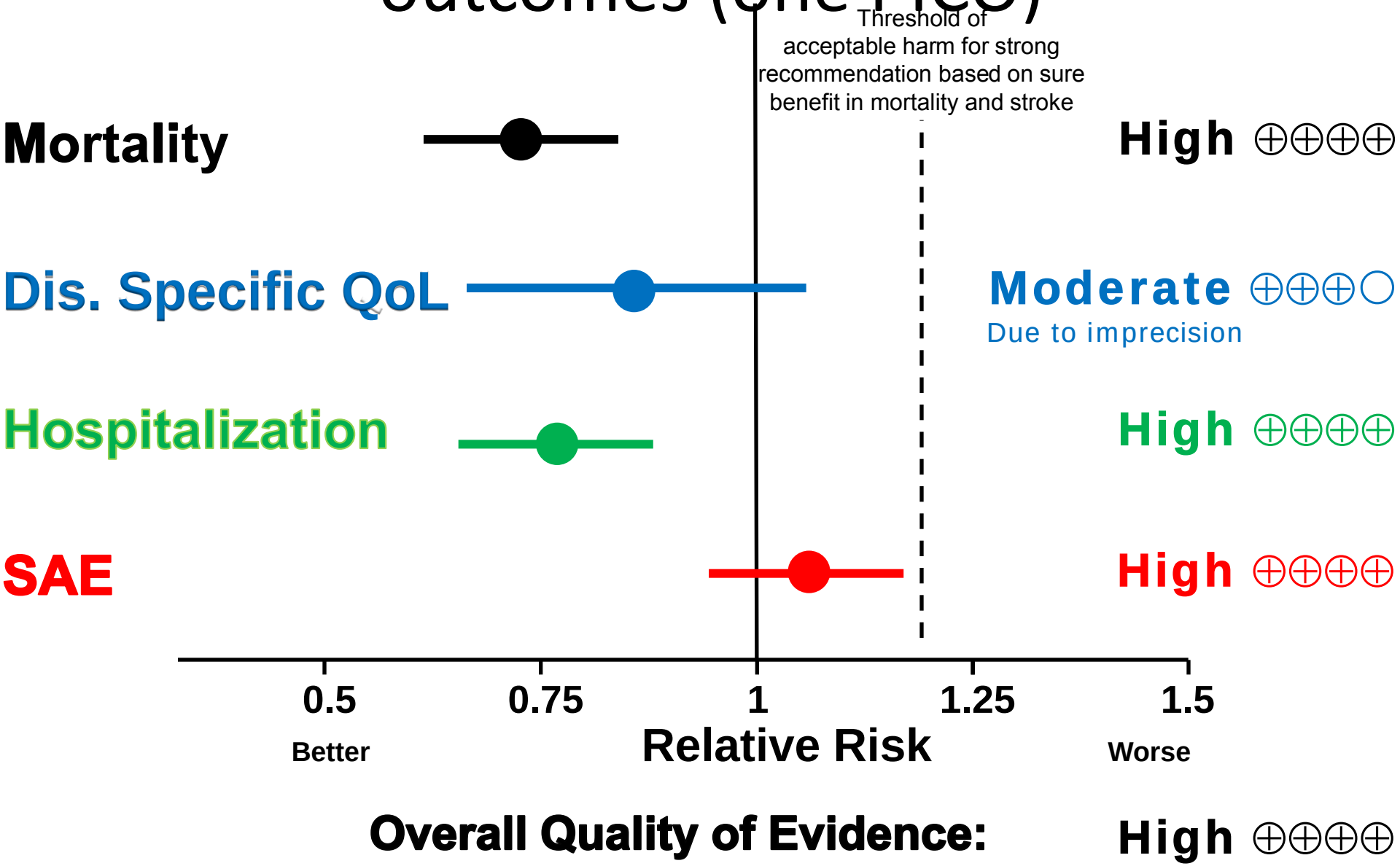
It's quite straightforward.

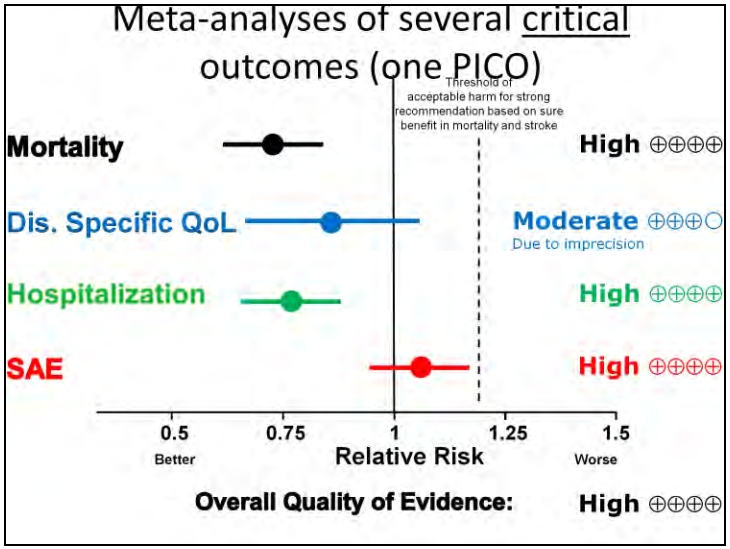
It's the lowest critical, so, yes.

So, the overall quality of evidence is not low because nausea, despite the fact that it is only low quality, it was rated as important and not critical.

So moderate.

Meta-analyses of several critical outcomes (one PICO)





Let's assume the following case.

Now, mortality is a critical -- So these are all critical outcomes now, all critical outcomes.

Mortality -- It's high-quality evidence that this intervention reduces mortality.

You were interested in disease-specific quality of life.

It was rated as a critical outcome, moderate due to precision.

Hospitalization was also high quality.

And serious adverse events was also high quality.

It was felt that there was ne'er enough confidence intervals to not downgrade.

So, what would the overall quality be here?

They're all critical.

Why would it be high?

So, if we were to apply the criteria that I just said, that it would be based on the lowest quality of the critical outcomes, it would be only moderate, right?

But either you did the reading or our common sense was similar to your common sense that it would be wrong to penalize this body of evidence.

So, you said three out of the four were high.

That could be one way of dealing with it.

Our way, or the way that we apply this criteria -- because it is really important for many of your questions, I believe -- is the following.

This outcome that would determine the lowest quality of evidence is actually going in the same direction, right?

And even having more information about it would not alter the recommendation that you would like to make because there are two critical outcomes that clearly go in one direction.

They cross the threshold for recommending an intervention against serious adverse events.

And it is very unlikely, apart from the fact that I just mentioned, that you would ever get more information into the specific quality of life.

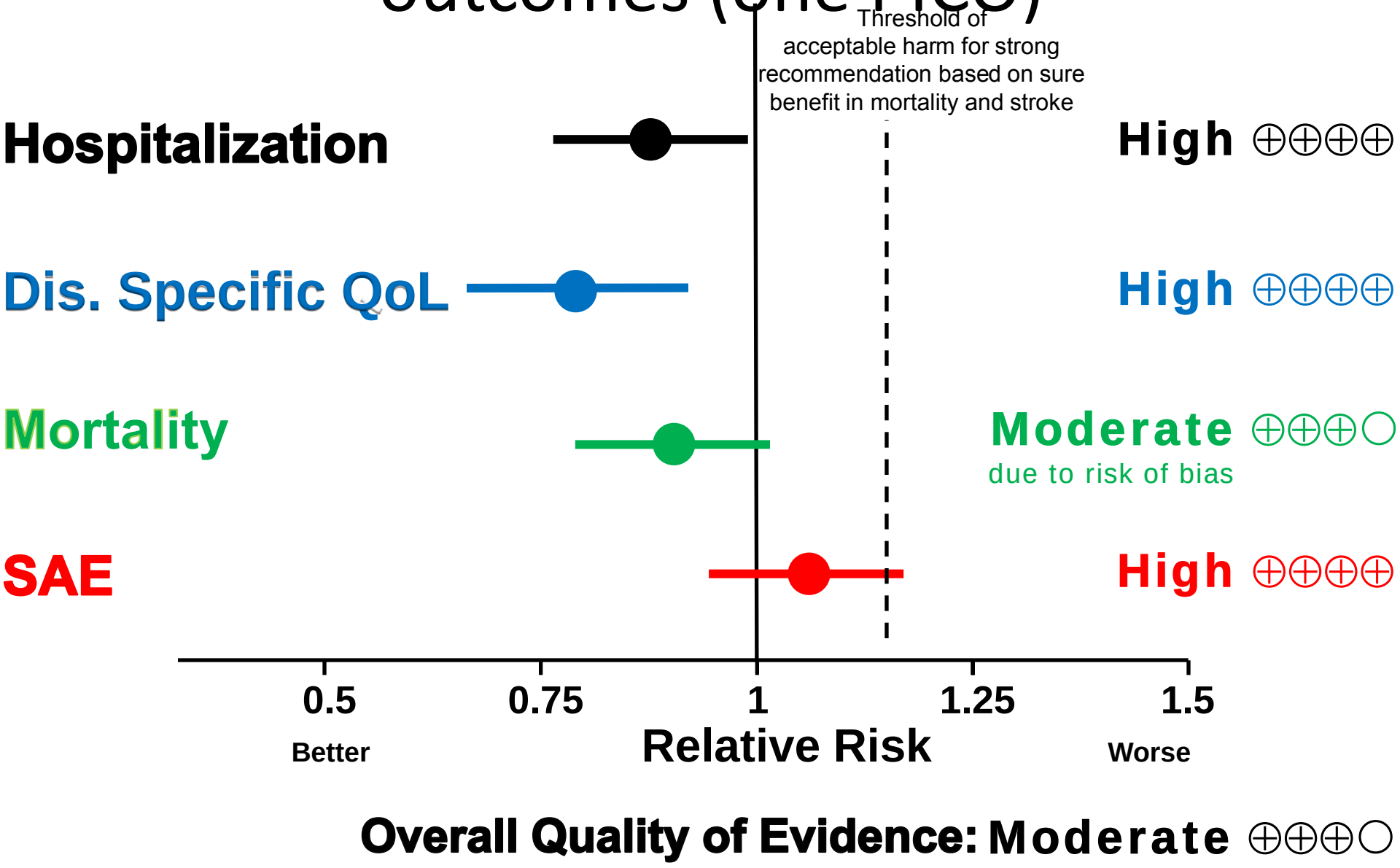
But the point is, it goes in the same direction with the other critical outcomes, and under those circumstances, we would not penalize the body of evidence and maintain a high quality rating.

And that, in particular, once again, if the threshold for the acceptable harm is crossed.

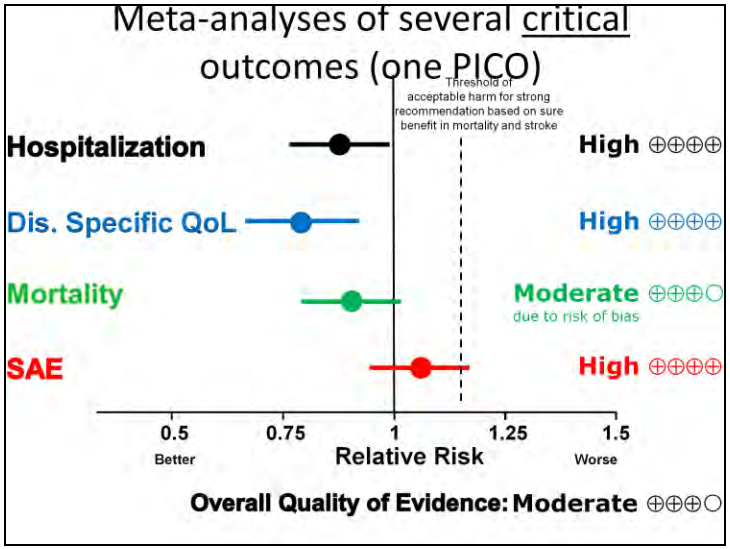
So where this is the threshold for where the serious adverse events should be falling into, considering the benefits that are obtained.

So the quality of the evidence would be high, rather than moderate.

Meta-analyses of several critical outcomes (one PICO)



Overall Quality of Evidence: Moderate ⊕⊕⊕◯



Last example -- All critical outcomes -- hospitalization is one outcome, disease-specific quality of life is another outcome, high mortality is moderate, and the serious adverse events are high.

And if you take all of this together -- You know, if you take these effects together and then look at how large a plausible increase in the risk of serious adverse events you would be willing to accept in order to recommend this -- If you consider that and if you consider that it wouldn't cross the threshold, that it would not be clearly on one side of the threshold, it means that you really do need additional information and that your overall confidence really should be reduced.

And under those circumstances, rightly so, the overall quality of the evidence would be moderate, based on the critical outcomes that you have here, the lowest critical outcome, in particular, because the threshold is not crossed.

So the overall quality is determined by the lowest critical outcome, except for the circumstances, the situation that I described there.

Interpretation of grades of evidence

- ⊕⊕⊕⊕/A/High: We are very confident that the true effect lies close to that of the estimate of the effect.
- ⊕⊕⊕◯/B/Moderate**:** We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- ⊕⊕◯◯/C/Low: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
- ⊕◯◯◯/D/Very low: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

Formulate question

Select outcomes

Rate importance

Outcomes across studies

Create evidence profile with GRADEpro

Rate quality of evidence for each outcome

McMaster University

Randomization increases initial quality

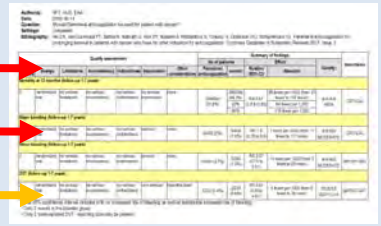
P
I
C
O

Outcome Critical

Outcome Critical

Outcome Important

Outcome Not important



Summary of findings & estimate of effect for each outcome

High
Moderate
Low
Very low

Grade down

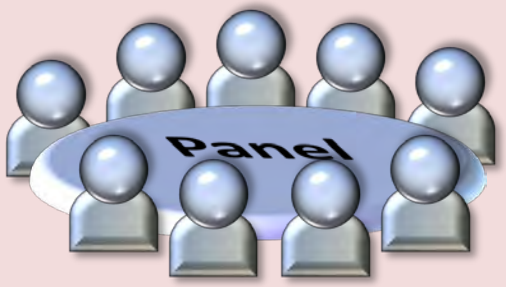
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

Grade up

1. Large effect
2. Dose response
3. Confounders

Systematic review

Guideline development



Grade overall quality of evidence across outcomes based on lowest quality of ***critical*** outcomes

Formulate recommendations:

- For or against (direction)
- Strong or conditional/weak (strength)

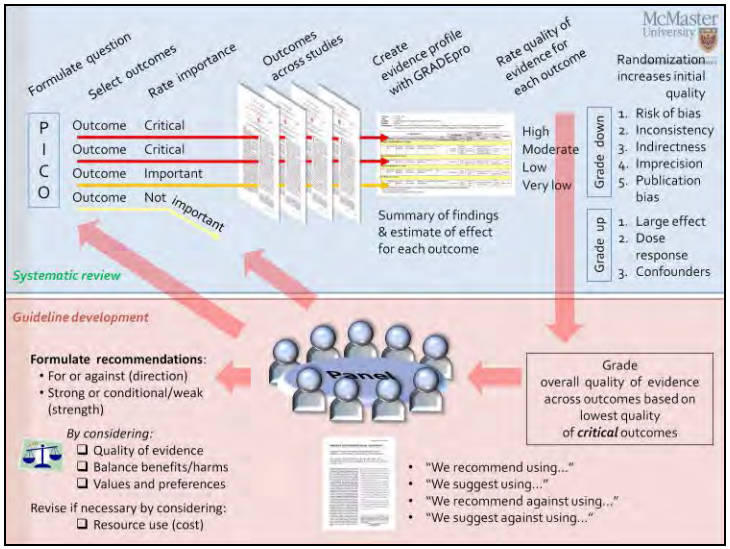
By considering:

- ❑ Quality of evidence
- ❑ Balance benefits/harms
- ❑ Values and preferences

Revise if necessary by considering:

- ❑ Resource use (cost)

- "We recommend using…"
- "We suggest using…"
- "We recommend against using…"
- "We suggest against using…"



This figure demonstrates the ideal process of integrating the GRADE approach into guideline development and the relation between systematic review conduct and guideline development. We will describe this process in an overview first and then describe selected single steps in more detail. It highlights that there is a requirement for a close relation between guideline panels, systematic reviews and those who assess the confidence in the estimates of effect (i.e. the quality of the evidence). It describes that guideline panels should be involved in the development of appropriate healthcare questions according to the PICO framework (reference article 3). The panel is involved in developing these outcomes and selecting the outcomes and in assessing their importance for decision making. This process requires close collaboration of the multidisciplinary panel. Outcomes that are considered critical and important are evaluated in a systematic review. Outcomes that are rated as not important do not have to be considered further. The novelty of the GRADE approach is that the outcomes are evaluated across studies rather than within studies. That is, a different body of evidence may contribute information to different outcomes that are being considered. When an evaluation of the outcomes across studies has taken place evidence profiles using software such as GRADEpro are developed the presentation of this information can either take place in typical evidence profiles or also in the Summary of Findings tables where a detailed assessment of the underlying confidence in an estimate of effect by outcome is then combined with an actual analysis of what the effects are. Those who review the evidence will then grade the confidence in the estimates of effect of a body of

evidence (i.e. the quality of evidence) for each outcome in four categories; high, moderate, low or very low on the basis of 8 factors that either increase or decrease the initial quality. Randomization is considered the best method to protect against bias and confounding and the initial quality of a body of evidence from randomized control trials usually starts as high quality, but there are 5 factors that lower the quality and, usually, for observational studies, 3 factors that increase the quality.

Once all outcomes that are critical for decision making have been evaluated an overall confidence in the estimate of effect to support a recommendation or an overall GRADE of the quality of evidence is assigned. The overall GRADE is based on the outcome with the lowest quality of evidence given that it is a critical outcome. This information is then provided back to the panel.

A guideline panel then needs to formulate a recommendation by considering the following 4 factors: the quality of evidence, the balance between benefits and down sides, values and preferences and resource use. A panel will then formulate recommendations in a clear and unambiguous way using standardized wording, such as using the term recommend for strong recommendations and suggest for conditional or weak recommendations or other terminology such as “should” and “may”. Guideline panels will express GRADE’s two directions of the recommendation either for or against an intervention or diagnostic test or strategy and the strength of this recommendation by either determining that it is a strong or a conditional recommendation. Other users of GRADE may use the evidence summarized according to the GRADE approach for health policy decisions.

Evidence Profiles/Summaries

McMaster University
Inspiring Innovation and Discovery

Table 1: Quality assessment							No of patients		Effect	Quality	Importance
No of studies	Design	Limitations[1]	Inconsistency[2]	Indirectness	Imprecision	Publication Bias[3]	ART use	No ART Use	Relative (95% CI)		
1. Cure (failure)											
9	observational studies	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	possible	33/72 (46%)	7/53 (13%)	HR 3.17[4] (1.46,6.90)	⊕⊕OO	CRITICAL
2. Prompt initiation of appropriate treatment											
see table 2											
3. Avoiding the acquisition or amplification of drug resistance											
9	observational studies	no serious limitations	no serious inconsistency	no serious indirectness	very serious[5]	possible			-	⊕OOO	CRITICAL
4. Death from TB											
10	observational studies	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	possible	34/124 (27%)	48/83 (58%)	HR 0.41[6] (0.26, 0.63)	⊕⊕⊕O[7]	CRITICAL
5a. Staying disease-free after treatment; sustaining a cure (relapse)											
Studies not identified to evaluate this outcome											
5b. Case holding so the TB patient remains adherent to treatment (default or treatment interruption due to non-adherence)											
9	observational studies	no serious limitations	no serious inconsistency	no serious indirectness	Serious[8]	possible	6/72 (8%)	9/53 (17%)	HR 0.48 (0.18, 1.31)	⊕OOO	CRITICAL
6. Population coverage or access to appropriate treatment of drug resistant TB- not measured											
Studies not identified to evaluate this outcome											
7a. Smear conversion during treatment											
4	observational studies	no serious limitations	no serious inconsistency	no serious indirectness	Serious[8]	possible	10/18 (56%)	13/20 (65%)	HR 1.11 (0.48, 2.57)	⊕OOO	CRITICAL
7a. Culture conversion during treatment											
5	observational studies	no serious limitations	no serious inconsistency	no serious indirectness	Serious[8]	possible	27/71 (38%)	17/50 (34%)	HR1.2 (0.65, 2.21)	⊕OOO	CRITICAL
7b. Accelerated detection of drug resistance											
not evaluated in the context of our question											
8. Avoid unnecessary MDR treatment											
Studies not identified to evaluate this outcome											
9. Population coverage or access to diagnosis of drug resistant TB											
not evaluated in the context of our question											
10. Prevention or interruption of transmission of DR TB to other people, including other patients, health care workers											
Studies not identified to evaluate this outcome											
11. Shortest possible duration of treatment											
Studies not identified to evaluate this outcome											
12. Avoiding toxicity and adverse reactions from TB drugs											

Agenda

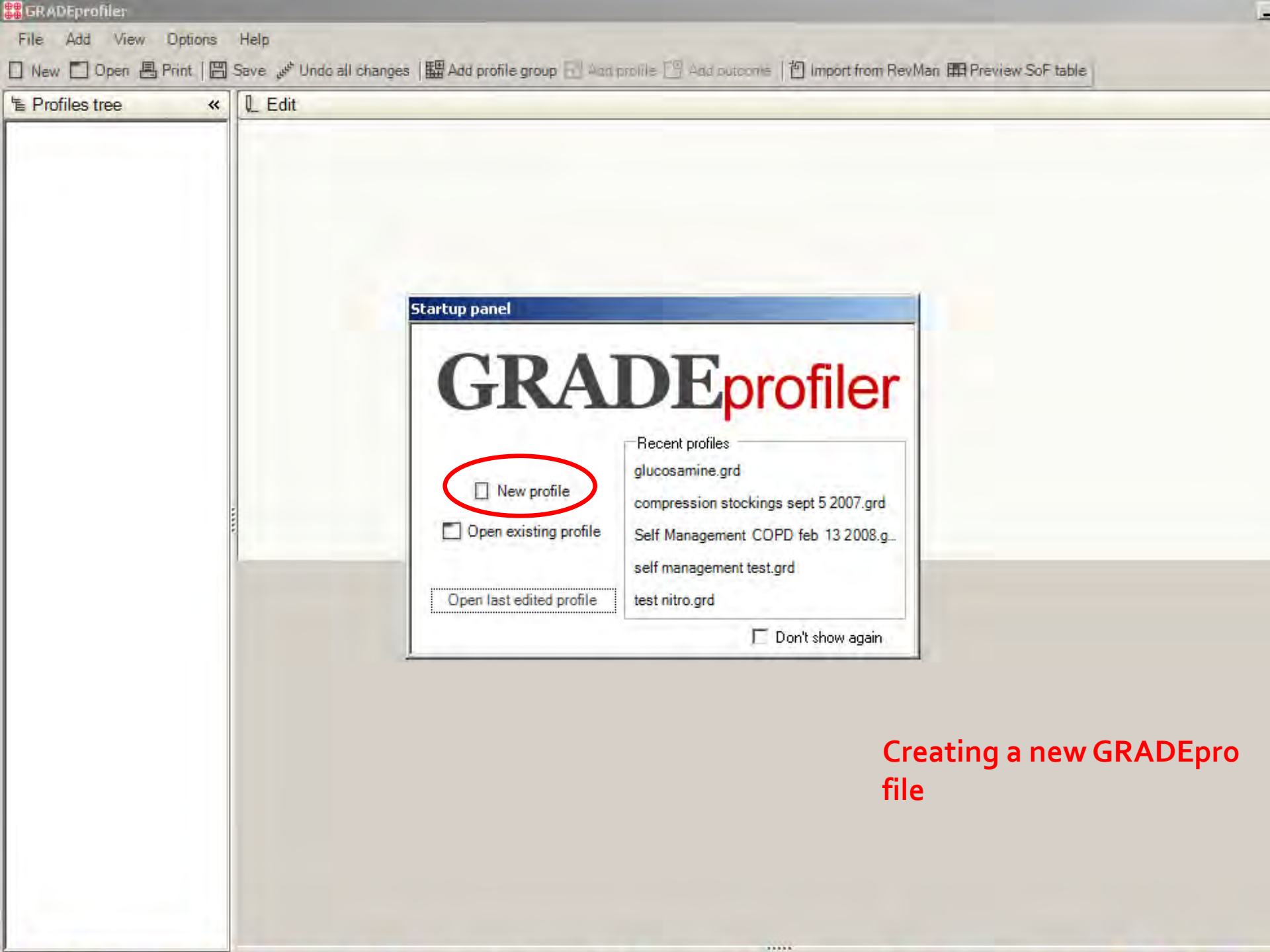
McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

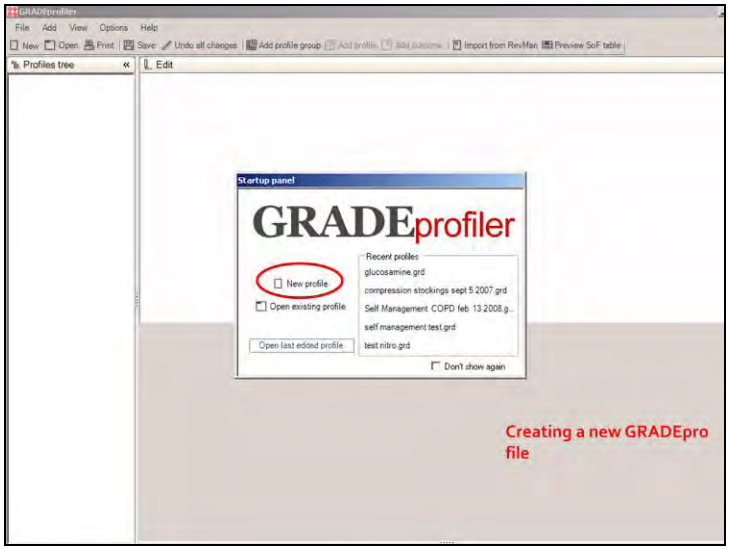
Agenda

McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**



Creating a new GRADEpro file



These evidence syntheses are typically prepared using the GRADE Profiler software, also called GRADE Pro that is freely available on the internet and that has functions that permit, for instance import from RAV Man, the systematic review and meta-analysis software that is produced by the Cochrane Collaboration, GRADE Pro is a simple to use software that allows the considered judgments that we have just described about the evidence and the production of GRADE evidence profiles as well as Summary of Findings tables.

GRADEprofiler help

Hide Back Forward Print

Contents Index Search Favorites

- GRADE Handbook
- Welcome to the GRADEprofiler
- What is new?
- Getting started
- Creating GRADE evidence profiles
  - Creating a group of profiles
  - Creating a profile
    - Defining a health care questic
    - Providing profile information
  - Creating an outcome
  - Grading quality of evidence
  - Summarizing the evidence
  - GRADE evidence profiles for obs
- Creating Summary of Findings Tables
- Managing evidence profiles
- Overview of the GRADE Approach
- Glossary of terms and concepts
- References to articles about GRADE
- Additional resources
- System requirements
- Technical support
- GRADE Working Group
- Licence agreement
- GRADEpro credits and acknowledge

Defining a Health Care Question

You should define a health care question whenever you create a new evidence profile.

We recommend that every comparison in a systematic review or a recommendation in guidelines is defined by a clear, explicit, and focused health care question. The clearer the question the easier it is to formulate an unambiguous conclusion or recommendation.

To formulate a question:

- choose question format from a drop-down menu
- specify the **intervention** being considered
- specify the alternative intervention (**comparison**)
- depending on the question format you have chosen, specify **patients or population** that the recommendation is intended for, or a **health problem** that the recommendation will address

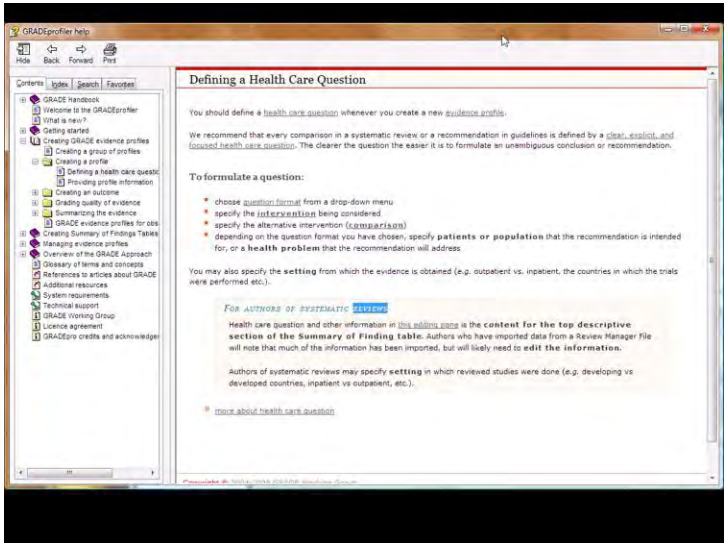
You may also specify the **setting** from which the evidence is obtained (*e.g.* outpatient vs. inpatient, the countries in which the trials were performed etc.).

FOR AUTHORS OF SYSTEMATIC REVIEWS

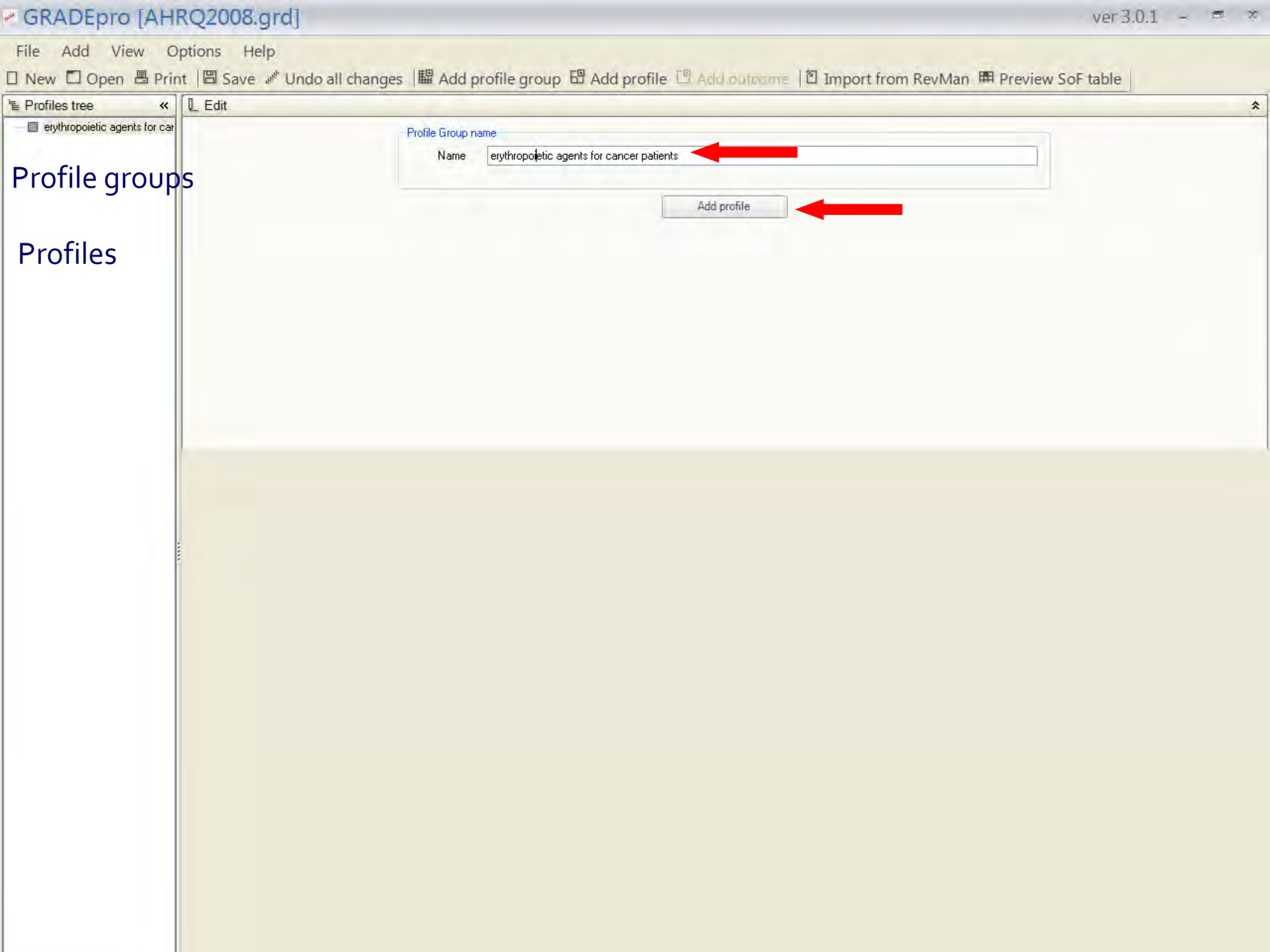
Health care question and other information in this editing pane is the **content for the top descriptive section of the Summary of Finding table**. Authors who have imported data from a Review Manager File will note that much of the information has been imported, but will likely need to **edit the information**.

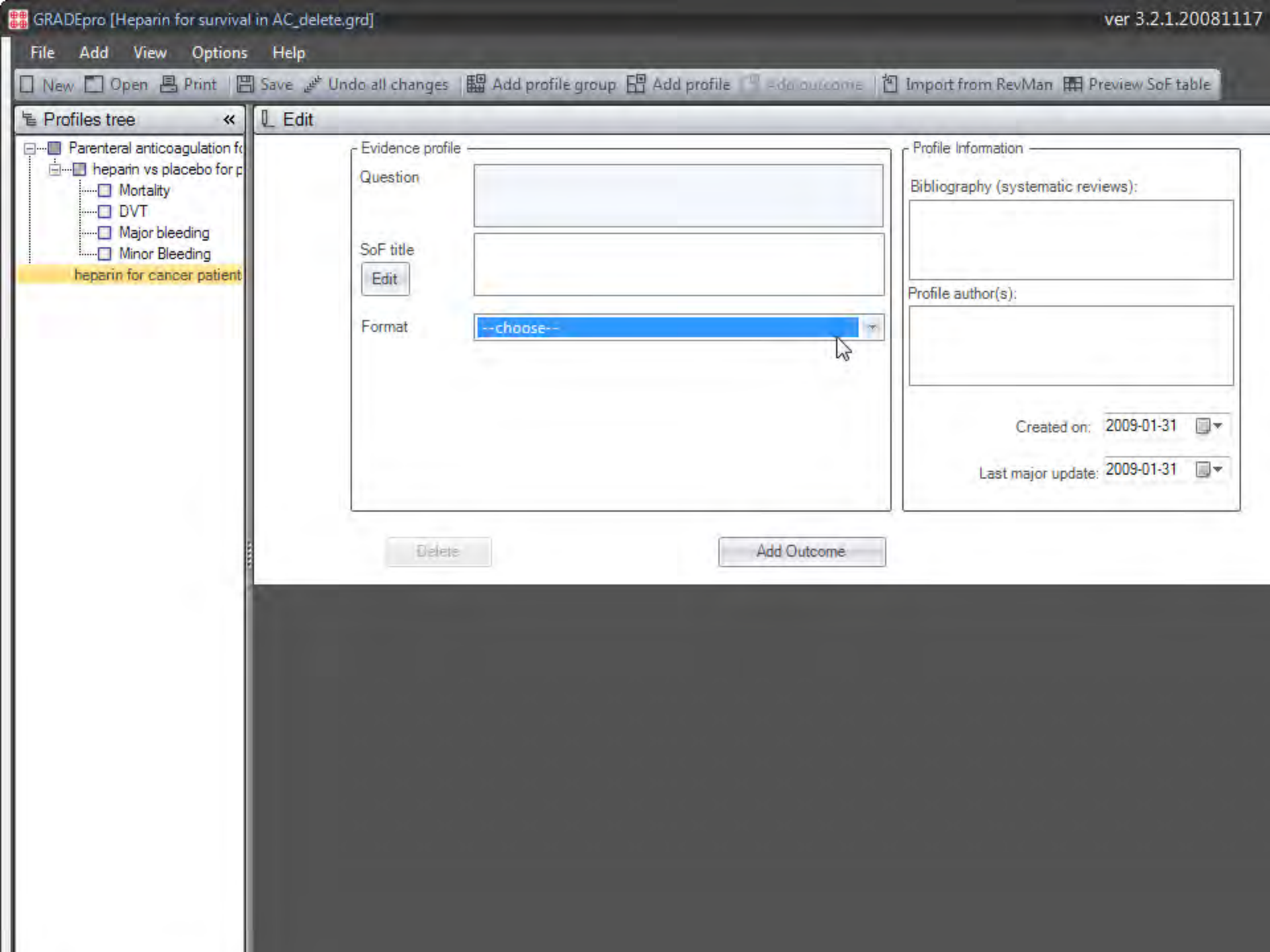
Authors of systematic reviews may specify **setting** in which reviewed studies were done (*e.g.* developing vs developed countries, inpatient vs outpatient, etc.).

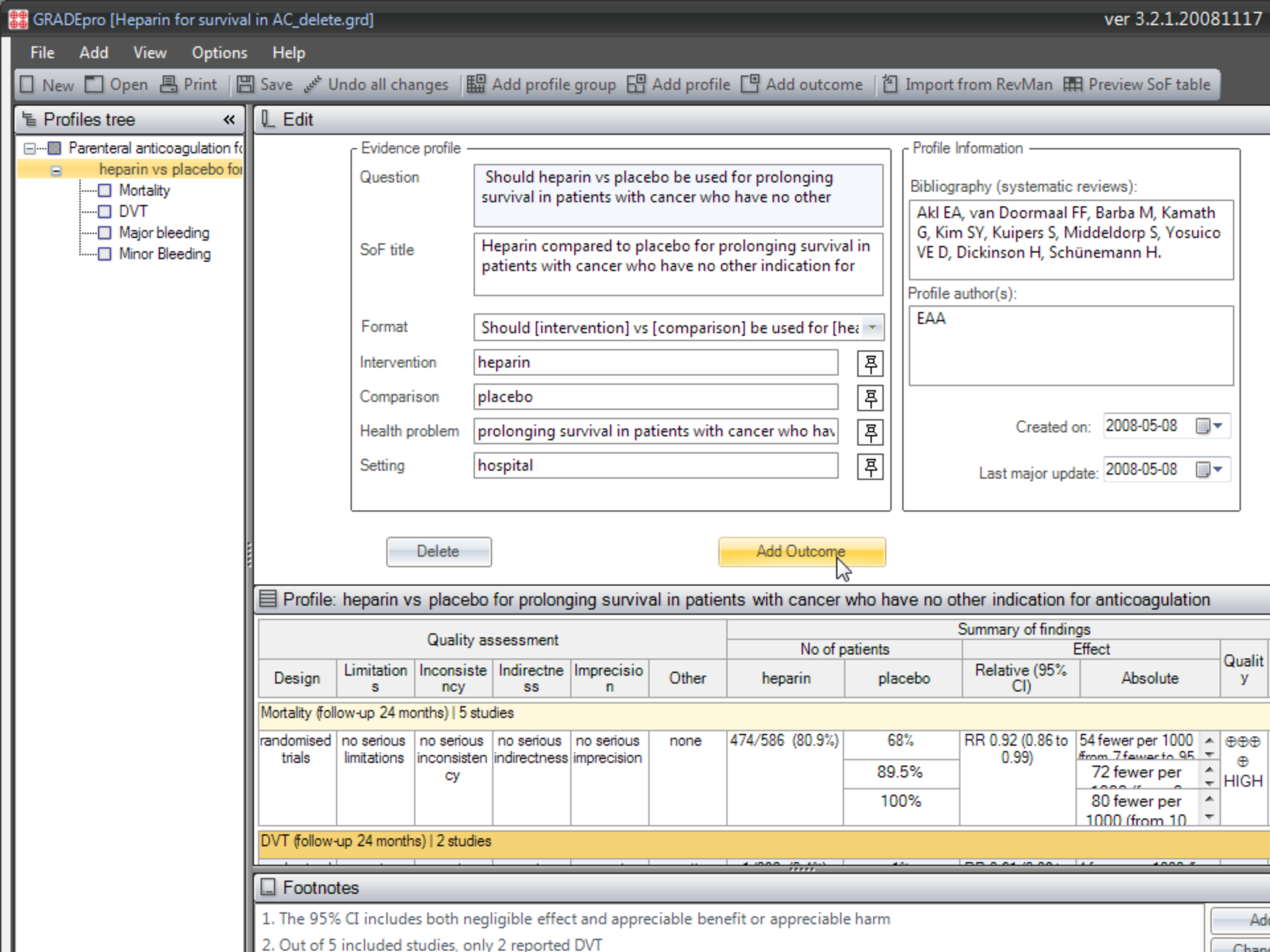
» more about health care question



One of the most important features of the GRADE Profiler software is that it includes a complete and very extensive handbook in the form of an electronic help file that allows understandings of the judgments that are made in GRADE and how evidence profiles and Summary of Findings tables are produced. In fact, this software is regularly updated with the newest developments in the GRADE Working Group and once again is the most up-to-date and comprehensive information about the GRADE approach.


GRADEpro [AHRQ2008.grd]
ver 3.0.1
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
erythropoietic agents for car
Edit
Profile Group name
Name
erythropoietic agents for cancer patients
Add profile
Profile groups
Profiles


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation fo
heparin vs placebo for p
Mortality
DVT
Major bleeding
Minor Bleeding
heparin for cancer patient
Edit
Evidence profile
Question
SoF title
Edit
Format
--choose--
Profile Information
Bibliography (systematic reviews):
Profile author(s):
Created on:
2009-01-31
Last major update:
2009-01-31
Delete
Add Outcome


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation fo
heparin vs placebo fo
Mortality
DVT
Major bleeding
Minor Bleeding
Edit
Evidence profile
Question
Should heparin vs placebo be used for prolonging survival in patients with cancer who have no other
SoF title
Heparin compared to placebo for prolonging survival in patients with cancer who have no other indication for
Format
Should [intervention] vs [comparison] be used for [hea
Intervention
heparin
Comparison
placebo
Health problem
prolonging survival in patients with cancer who hav
Setting
hospital
Profile Information
Bibliography (systematic reviews):
Akl EA, van Doormaal FF, Barba M, Kamath G, Kim SY, Kuipers S, Middeldorp S, Yosuico VE D, Dickinson H, Schünemann H.
Profile author(s):
EAA
Created on:
2008-05-08
Last major update:
2008-05-08
Delete
Add Outcome
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Quality assessment
Summary of findings
No of patients
Effect
Design
Limitations
Inconsistency
Indirectness
Imprecision
Other
heparin
placebo
Relative (95% CI)
Absolute
Quality
Mortality (follow-up 24 months) | 5 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
no serious imprecision
none
474/586 (80.9%)
68%
89.5%
100%
RR 0.92 (0.86 to 0.99)
54 fewer per 1000
72 fewer per
80 fewer per 1000 (from 10
⊕⊕⊕⊕ HIGH
DVT (follow-up 24 months) | 2 studies
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT

GRADEpro [Heparin for survival in AC_delete.grd] ver 3.2.1.20081117

File Add View Options Help

New Open Print Save Undo all changes Add profile group Add profile Add outcome Import from RevMan Preview SoF table

Profiles tree

- Parenteral anticoagulation fo
  - heparin vs placebo for
    - Mortality
    - DVT
    - Major bleeding
    - Minor Bleeding

Edit

Evidence profile

Question: Should heparin vs placebo be used for prolonging survival in patients with cancer who have no other

SoF title: Heparin compared to placebo for prolonging survival in patients with cancer who have no other indication for

Format: Should [intervention] vs [comparison] be used for [hea

Intervention: heparin

Comparison: placebo

Health problem: prolonging survival in patients with cancer who hav

Setting: hospital

Profile Information

Bibliography (systematic reviews):

Akl EA, van Doormaal FF, Barba M, Kamath G, Kim SY, Kuipers S, Middeldorp S, Yosuico VE D, Dickinson H, Schünemann H.

Profile author(s):

EAA

Created on: 2008-05-08

Last major update: 2008-05-08

Delete Add Outcome

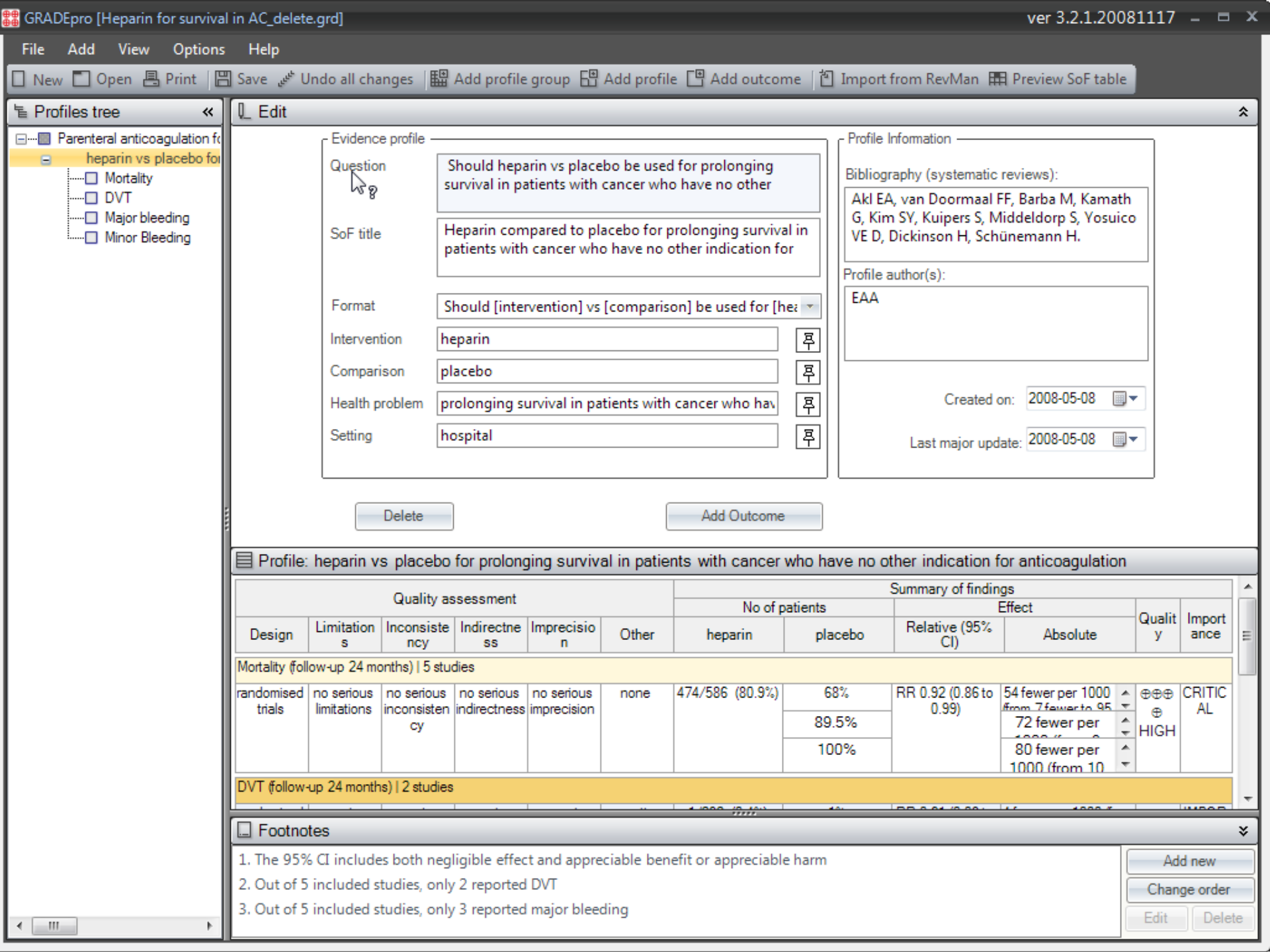
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation

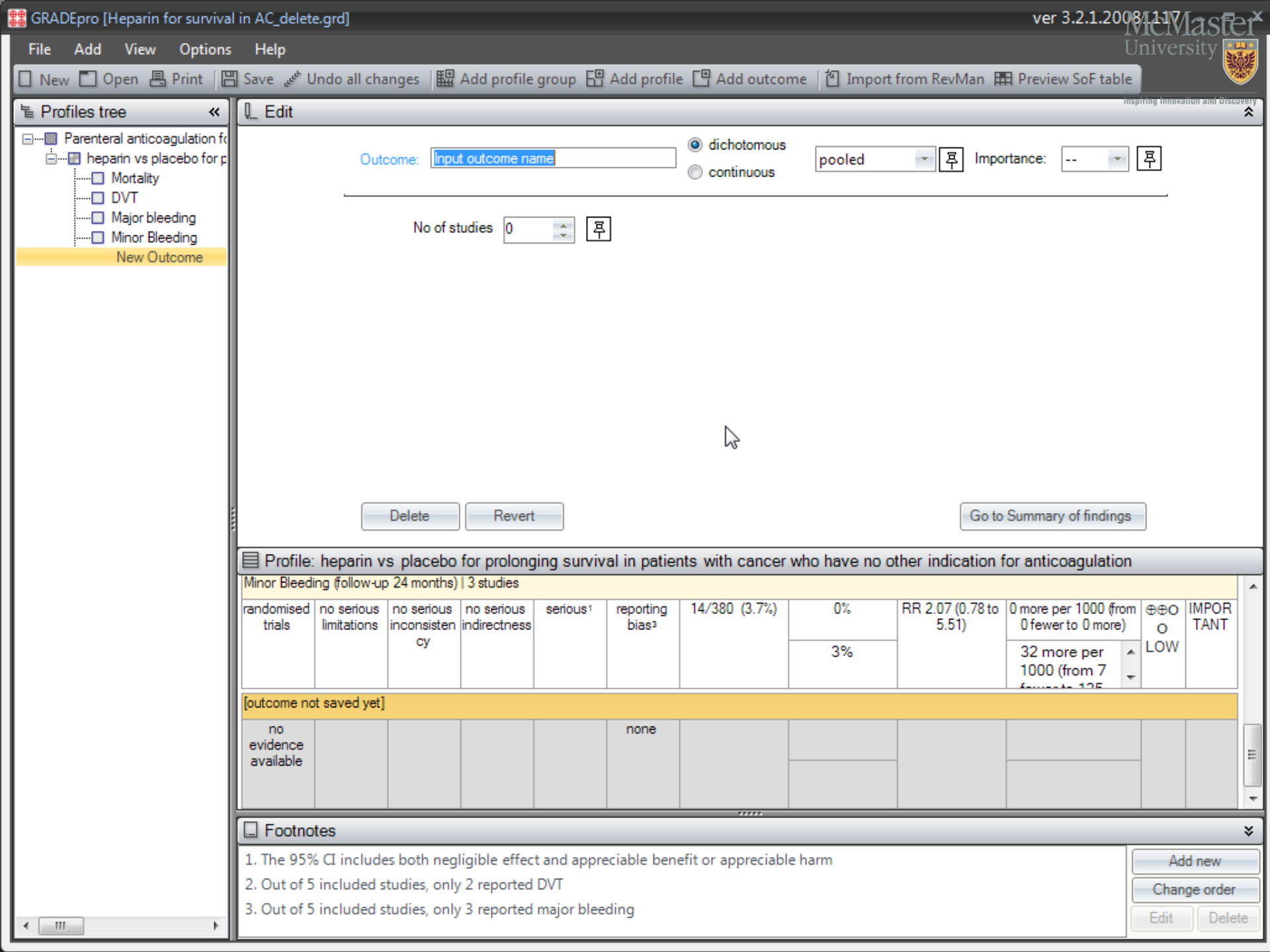
Quality assessment						Summary of findings					
						No of patients		Effect		Quality	Importance
Design	Limitations	Inconsistency	Indirectness	Imprecision	Other	heparin	placebo	Relative (95% CI)	Absolute		
Mortality (follow-up 24 months) | 5 studies											
randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	474/586 (80.9%)	68%	RR 0.92 (0.86 to 0.99)	54 fewer per 1000 (from 7 fewer to 95	⊕⊕⊕⊕ HIGH	CRITICAL
							89.5%		72 fewer per		
							100%		80 fewer per 1000 (from 10		
DVT (follow-up 24 months) | 2 studies											

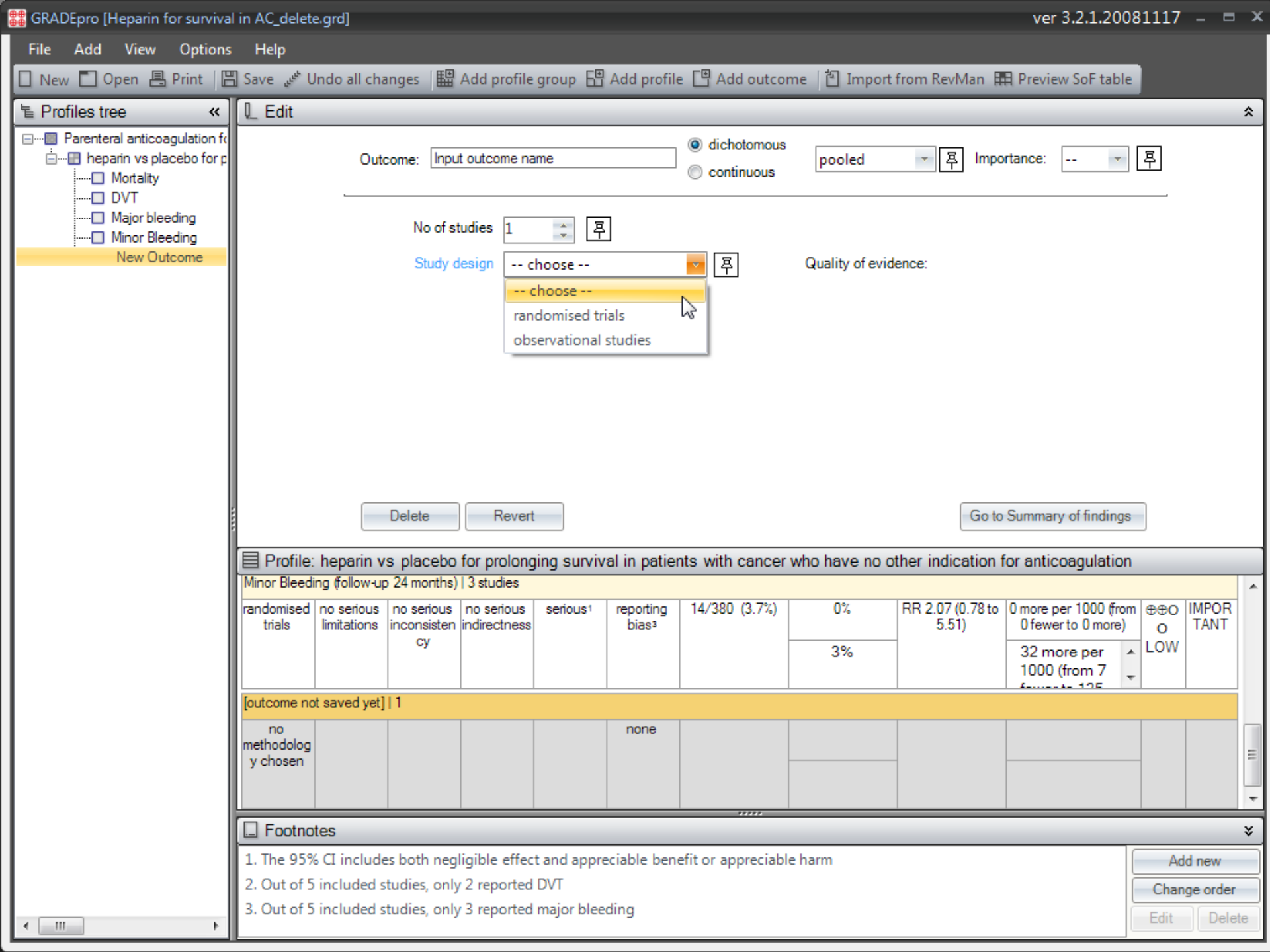
Footnotes

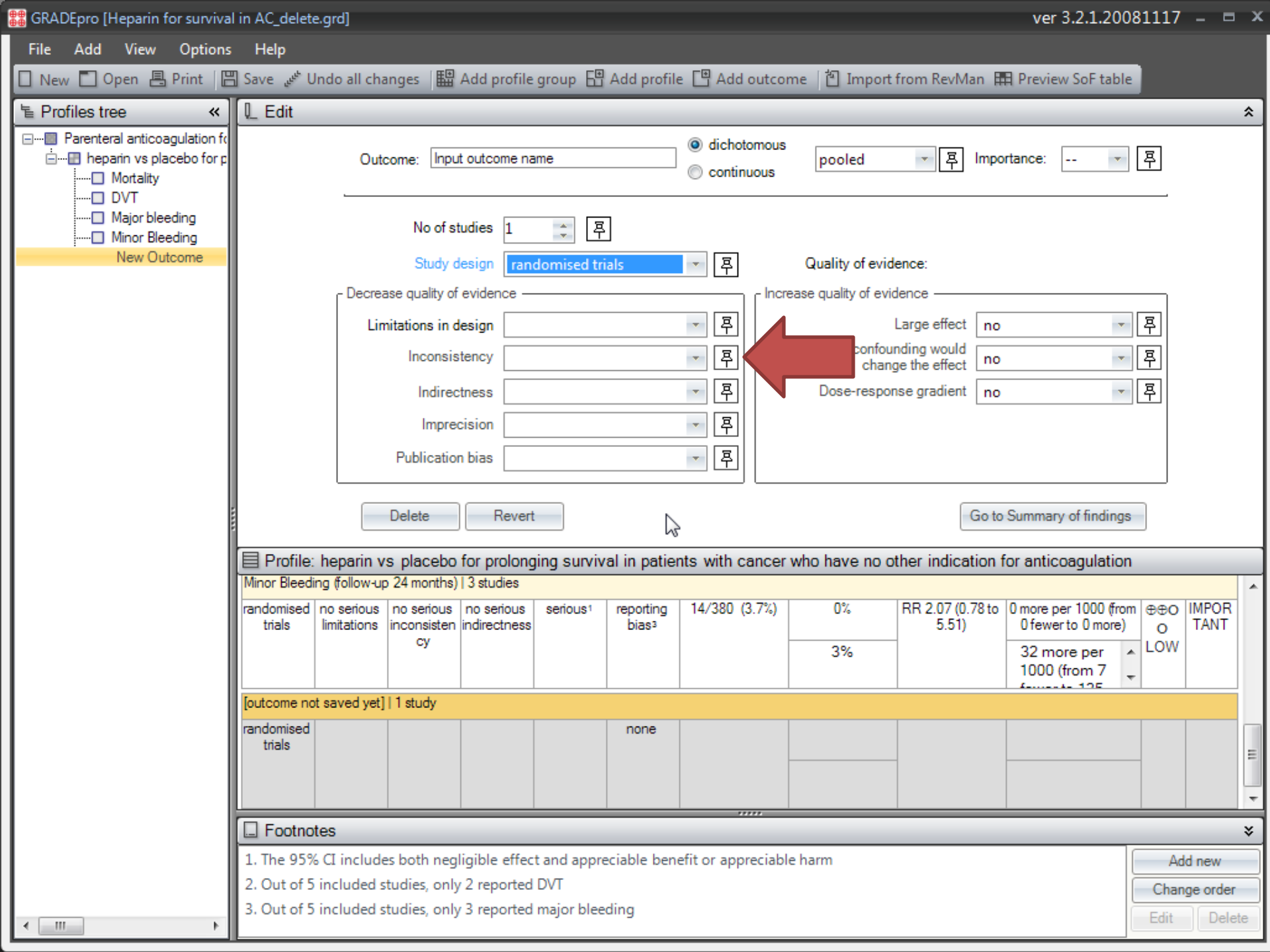
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding

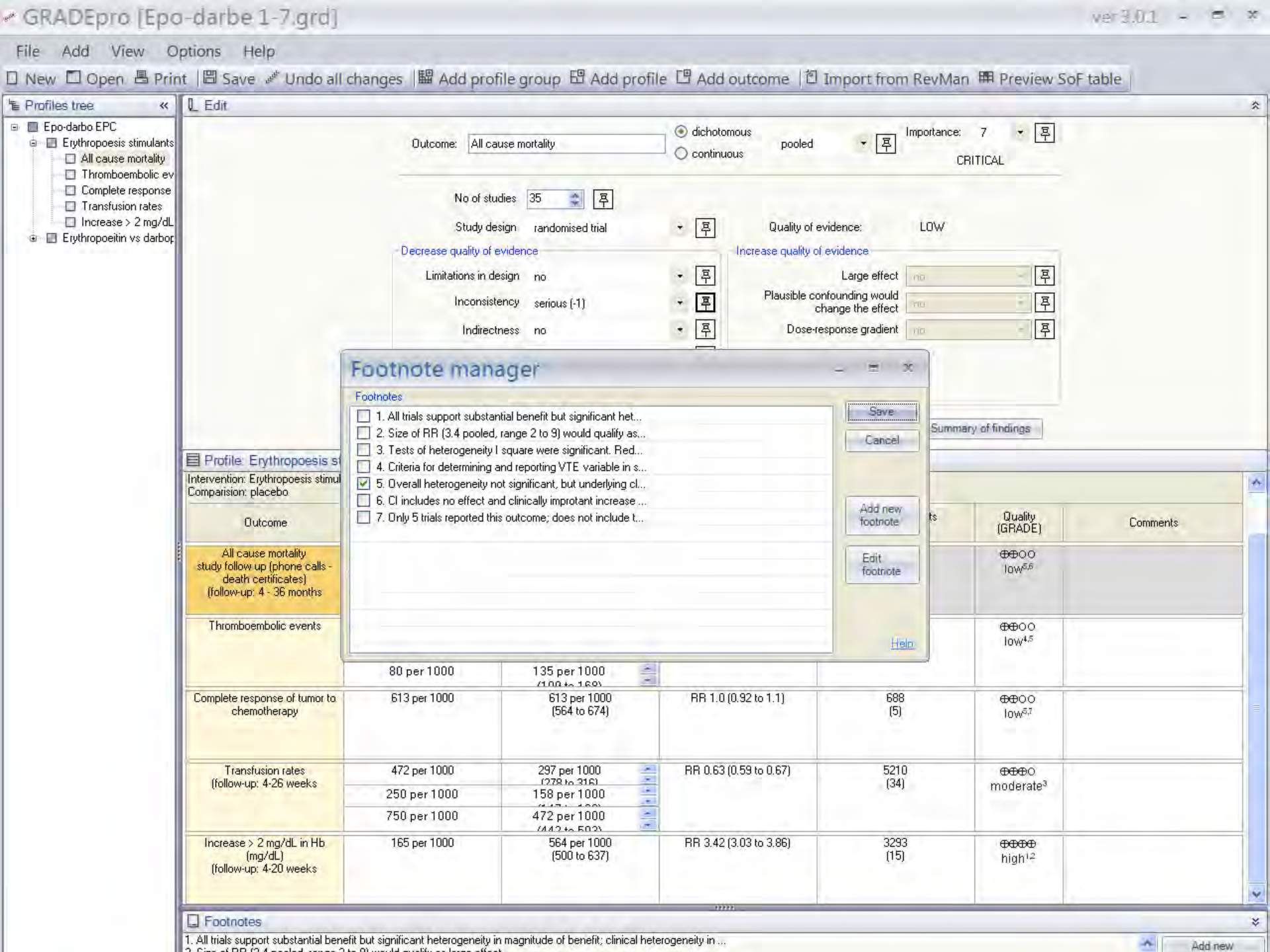
Add new Change order Edit Delete

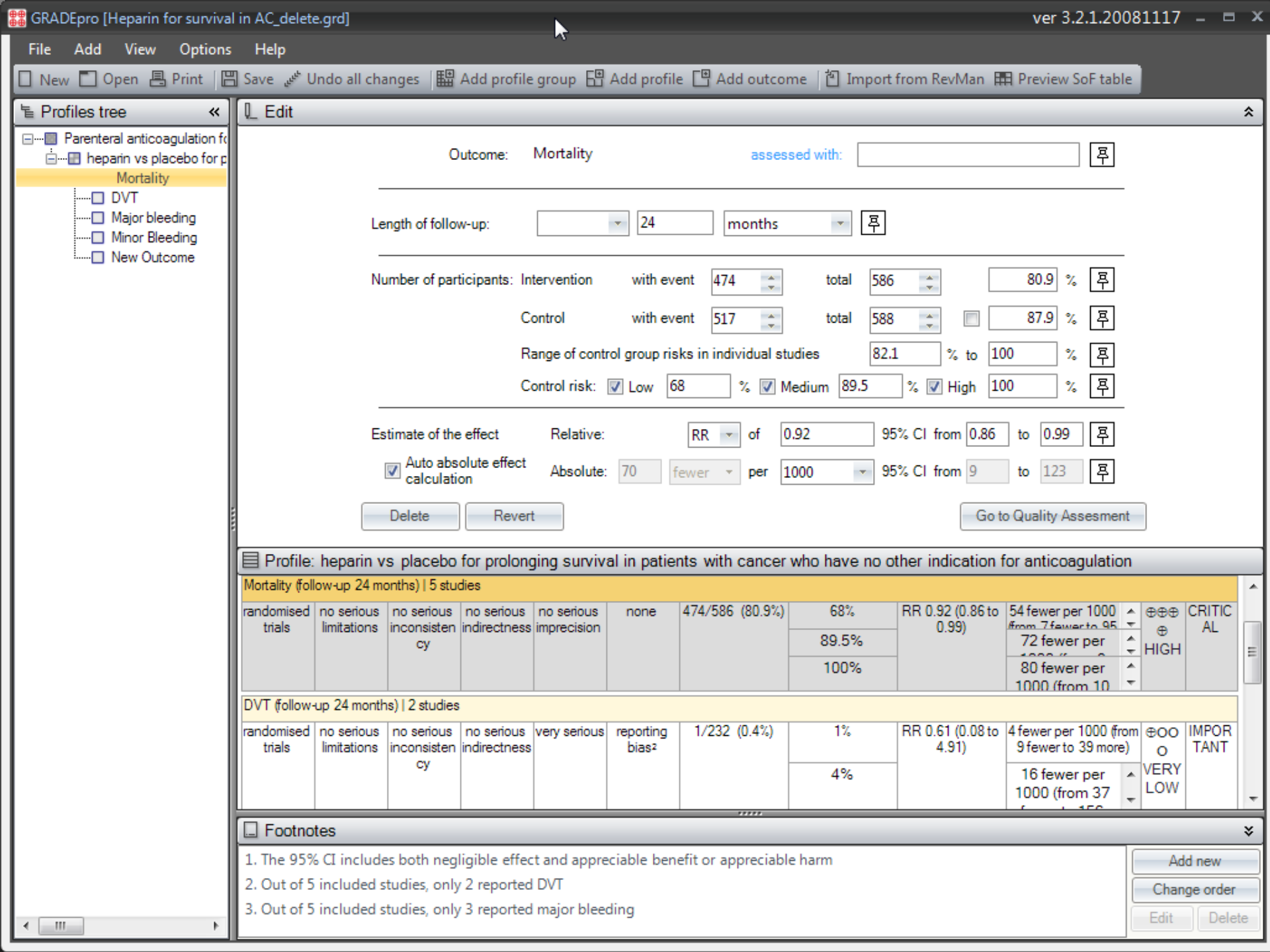

GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation fo
heparin vs placebo fo
Mortality
DVT
Major bleeding
Minor Bleeding
Edit
Evidence profile
Question
Should heparin vs placebo be used for prolonging survival in patients with cancer who have no other
SoF title
Heparin compared to placebo for prolonging survival in patients with cancer who have no other indication for
Format
Should [intervention] vs [comparison] be used for [hea
Intervention
heparin
Comparison
placebo
Health problem
prolonging survival in patients with cancer who hav
Setting
hospital
Profile Information
Bibliography (systematic reviews):
Akl EA, van Doormaal FF, Barba M, Kamath G, Kim SY, Kuipers S, Middeldorp S, Yosuico VE D, Dickinson H, Schünemann H.
Profile author(s):
EAA
Created on:
2008-05-08
Last major update:
2008-05-08
Delete
Add Outcome
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Quality assessment
Summary of findings
No of patients
Effect
Design
Limitations
Inconsistency
Indirectness
Imprecision
Other
heparin
placebo
Relative (95% CI)
Absolute
Quality
Importance
Mortality (follow-up 24 months) | 5 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
no serious imprecision
none
474/586 (80.9%)
68%
89.5%
100%
RR 0.92 (0.86 to 0.99)
54 fewer per 1000
72 fewer per
80 fewer per 1000 (from 10
⊕⊕⊕⊕ HIGH
CRITICAL
DVT (follow-up 24 months) | 2 studies
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding
Add new
Change order
Edit
Delete

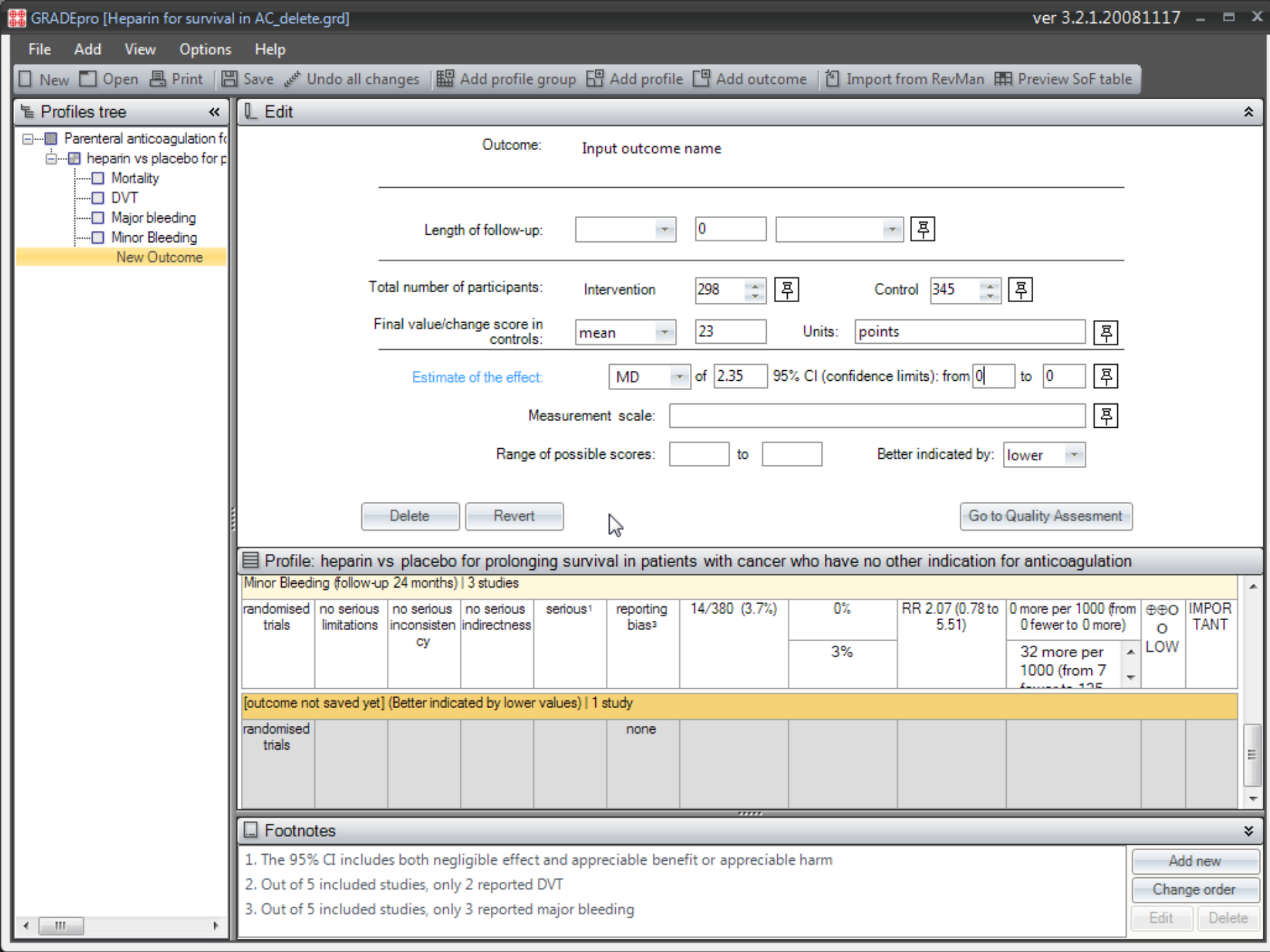

GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.2008.1117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation fo
heparin vs placebo for p
Mortality
DVT
Major bleeding
Minor Bleeding
New Outcome
Edit
Outcome:
Input outcome name
dichotomous
continuous
pooled
Importance:
--
No of studies
0
Delete
Revert
Go to Summary of findings
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Minor Bleeding (follow-up 24 months) | 3 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
serious1
reporting bias3
14/380 (3.7%)
0%
3%
RR 2.07 (0.78 to 5.51)
0 more per 1000 (from 0 fewer to 0 more)
32 more per 1000 (from 7
⊕⊕OO LOW
IMPORTANT
[outcome not saved yet]
no evidence available
none
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding
Add new
Change order
Edit
Delete

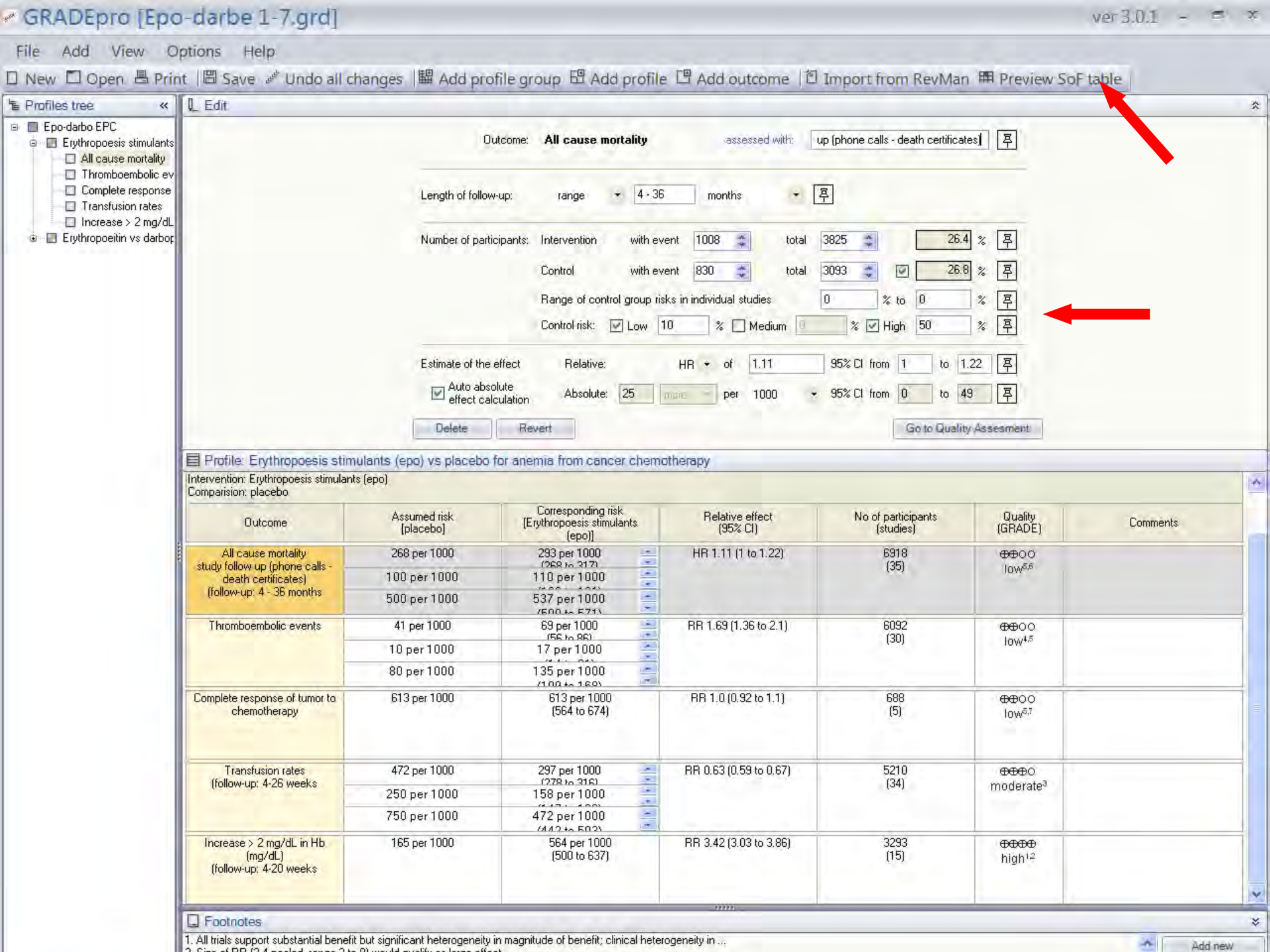

GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation f
heparin vs placebo for p
Mortality
DVT
Major bleeding
Minor Bleeding
New Outcome
Edit
Outcome:
Input outcome name
dichotomous
continuous
pooled
Importance:
--
No of studies
1
Study design
-- choose --
Quality of evidence:
-- choose --
randomised trials
observational studies
Delete
Revert
Go to Summary of findings
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Minor Bleeding (follow-up 24 months) | 3 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
serious1
reporting bias3
14/380 (3.7%)
0%
3%
RR 2.07 (0.78 to 5.51)
0 more per 1000 (from 0 fewer to 0 more)
32 more per 1000 (from 7
⊕⊕OO LOW
IMPORTANT
[outcome not saved yet] | 1
no methodology chosen
none
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding
Add new
Change order
Edit
Delete


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation f
heparin vs placebo for p
Mortality
DVT
Major bleeding
Minor Bleeding
New Outcome
Edit
Outcome:
Input outcome name
dichotomous
continuous
pooled
Importance:
--
No of studies
1
Study design
randomised trials
Quality of evidence:
Decrease quality of evidence
Limitations in design
Inconsistency
Indirectness
Imprecision
Publication bias
Increase quality of evidence
Large effect
no
confounding would change the effect
no
Dose-response gradient
no
Delete
Revert
Go to Summary of findings
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Minor Bleeding (follow-up 24 months) | 3 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
serious[1]
reporting bias[3]
14/380 (3.7%)
0%
3%
RR 2.07 (0.78 to 5.51)
0 more per 1000 (from 0 fewer to 0 more)
32 more per 1000 (from 7
⊕⊕OO LOW
IMPORTANT
[outcome not saved yet] | 1 study
randomised trials
none
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding
Add new
Change order
Edit
Delete


GRADEpro [Epo-darbe 1-7.grd]
ver 3.0.1
File Add View Options Help
New Open Print Save Undo all changes Add profile group Add profile Add outcome Import from RevMan Preview SoF table
Profiles tree
Epo-darbo EPC
Erythropoesis stimulants
All cause mortality
Thromboembolic ev
Complete response
Transfusion rates
Increase > 2 mg/dL
Erythropoeitin vs darbop
Edit
Outcome: All cause mortality
dichotomous
continuous
pooled
Importance: 7
CRITICAL
No of studies 35
Study design randomised trial
Quality of evidence: LOW
Decrease quality of evidence
Limitations in design no
Inconsistency serious (-1)
Indirectness no
Increase quality of evidence
Large effect no
Plausible confounding would change the effect no
Dose-response gradient no
Footnote manager
Footnotes
1. All trials support substantial benefit but significant het...
2. Size of RR (3.4 pooled, range 2 to 9) would qualify as...
3. Tests of heterogeneity I square were significant. Red...
4. Criteria for determining and reporting VTE variable in s...
5. Overall heterogeneity not significant, but underlying cl...
6. CI includes no effect and clinically improtant increase ...
7. Only 5 trials reported this outcome; does not include t...
Save
Cancel
Add new footnote
Edit footnote
Help
Summary of findings
Profile: Erythropoesis st
Intervention: Erythropoesis stimul
Comparision: placebo
Outcome
Quality (GRADE)
Comments
All cause mortality study follow up (phone calls - death certificates) (follow-up: 4 - 36 months
low5,6
Thromboembolic events
low4,5
80 per 1000
135 per 1000
Complete response of tumor to chemotherapy
613 per 1000
613 per 1000 (564 to 674)
RR 1.0 (0.92 to 1.1)
688 (5)
low5,7
Transfusion rates (follow-up: 4-26 weeks
472 per 1000
297 per 1000
RR 0.63 (0.59 to 0.67)
5210 (34)
moderate3
250 per 1000
158 per 1000
750 per 1000
472 per 1000
Increase > 2 mg/dL in Hb (mg/dL) (follow-up: 4-20 weeks
165 per 1000
564 per 1000 (500 to 637)
RR 3.42 (3.03 to 3.86)
3293 (15)
high1,2
Footnotes
1. All trials support substantial benefit but significant heterogeneity in magnitude of benefit; clinical heterogeneity in ...
Add new


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation f
heparin vs placebo for p
Mortality
DVT
Major bleeding
Minor Bleeding
New Outcome
Edit
Outcome: Mortality
assessed with:
Length of follow-up:
24
months
Number of participants: Intervention with event 474 total 586 80.9 %
Control with event 517 total 588 87.9 %
Range of control group risks in individual studies 82.1 % to 100 %
Control risk: Low 68 % Medium 89.5 % High 100 %
Estimate of the effect Relative: RR of 0.92 95% CI from 0.86 to 0.99
Auto absolute effect calculation Absolute: 70 fewer per 1000 95% CI from 9 to 123
Delete
Revert
Go to Quality Assesment
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Mortality (follow-up 24 months) | 5 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
no serious imprecision
none
474/586 (80.9%)
68%
89.5%
100%
RR 0.92 (0.86 to 0.99)
54 fewer per 1000
72 fewer per
80 fewer per 1000 (from 10
⊕⊕⊕⊕ HIGH
CRITICAL
DVT (follow-up 24 months) | 2 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
very serious
reporting bias[2]
1/232 (0.4%)
1%
4%
RR 0.61 (0.08 to 4.91)
4 fewer per 1000 (from 9 fewer to 39 more)
16 fewer per 1000 (from 37
⊕OOO VERY LOW
IMPORTANT
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding
Add new
Change order
Edit
Delete


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New
Open
Print
Save
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoagulation fo
heparin vs placebo for p
Mortality
DVT
Major bleeding
Minor Bleeding
New Outcome
Edit
Outcome:
Input outcome name
Length of follow-up:
0
Total number of participants:
Intervention
298
Control
345
Final value/change score in controls:
mean
23
Units:
points
Estimate of the effect:
MD
of
2.35
95% CI (confidence limits): from
0
to
0
Measurement scale:
Range of possible scores:
to
Better indicated by:
lower
Delete
Revert
Go to Quality Assesment
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
Minor Bleeding (follow-up 24 months) | 3 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
serious1
reporting bias3
14/380 (3.7%)
0%
3%
RR 2.07 (0.78 to 5.51)
0 more per 1000 (from 0 fewer to 0 more)
32 more per 1000 (from 7
⊕⊕OO LOW
IMPORTANT
[outcome not saved yet] (Better indicated by lower values) | 1 study
randomised trials
none
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT
3. Out of 5 included studies, only 3 reported major bleeding
Add new
Change order
Edit
Delete


GRADEpro [Epo-darbe 1-7.grd]
ver 3.0.1
File Add View Options Help
New Open Print Save Undo all changes Add profile group Add profile Add outcome Import from RevMan Preview SoF table
Profiles tree
Epo-darbo EPC
Erythropoesis stimulants
All cause mortality
Thromboembolic ev
Complete response
Transfusion rates
Increase > 2 mg/dL
Erythropoeitin vs darbop
Edit
Outcome: All cause mortality
assessed with: up (phone calls - death certificates)
Length of follow-up: range 4 - 36 months
Number of participants: Intervention with event 1008 total 3825 26.4 %
Control with event 830 total 3093 26.8 %
Range of control group risks in individual studies 0 % to 0 %
Control risk: Low 10 % Medium % High 50 %
Estimate of the effect Relative: HR of 1.11 95% CI from 1 to 1.22
Auto absolute effect calculation Absolute: 25 more per 1000 95% CI from 0 to 49
Delete Revert
Go to Quality Assesment
Profile: Erythropoesis stimulants (epo) vs placebo for anemia from cancer chemotherapy
Intervention: Erythropoesis stimulants (epo)
Comparision: placebo
Outcome
Assumed risk [placebo]
Corresponding risk [Erythropoesis stimulants (epo)]
Relative effect (95% CI)
No of participants (studies)
Quality (GRADE)
Comments
All cause mortality study follow up (phone calls - death certificates) (follow-up: 4 - 36 months
268 per 1000
293 per 1000 (268 to 317)
HR 1.11 (1 to 1.22)
6918 (35)
⊕⊕OO low[5,6]
100 per 1000
110 per 1000
500 per 1000
537 per 1000 (500 to 571)
Thromboembolic events
41 per 1000
69 per 1000 (56 to 86)
RR 1.69 (1.36 to 2.1)
6092 (30)
⊕⊕OO low[4,5]
10 per 1000
17 per 1000
80 per 1000
135 per 1000 (100 to 169)
Complete response of tumor to chemotherapy
613 per 1000
613 per 1000 (564 to 674)
RR 1.0 (0.92 to 1.1)
688 (5)
⊕⊕OO low[5,7]
Transfusion rates (follow-up: 4-26 weeks
472 per 1000
297 per 1000 (278 to 316)
RR 0.63 (0.59 to 0.67)
5210 (34)
⊕⊕⊕O moderate[3]
250 per 1000
158 per 1000
750 per 1000
472 per 1000 (442 to 503)
Increase > 2 mg/dL in Hb (mg/dL) (follow-up: 4-20 weeks
165 per 1000
564 per 1000 (500 to 637)
RR 3.42 (3.03 to 3.86)
3293 (15)
⊕⊕⊕⊕ high[1,2]
Footnotes
1. All trials support substantial benefit but significant heterogeneity in magnitude of benefit; clinical heterogeneity in ...
Add new

Export as Html

Selected Profile(s)

Erythropoesis stimulants (epo) vs placebo for anemia fr
Erythropoeitin vs darbopoeitin for anemia from cancer c

Save
Close

Select format

- (•) GRADE evidence table
- ( ) Summary of findings table
- ( ) Overview of findings table

Select Outcomes
Histogram/Forest
Help

Author(s): DA, YFY
Date: 2008-01-07
Question: Should Erythropoesis stimulants (epo) vs placebo be used for anemia from cancer chemotherapy?
Settings: Outpatient cancer treatment
Bibliography: Effective health care #3 (AHRQ)

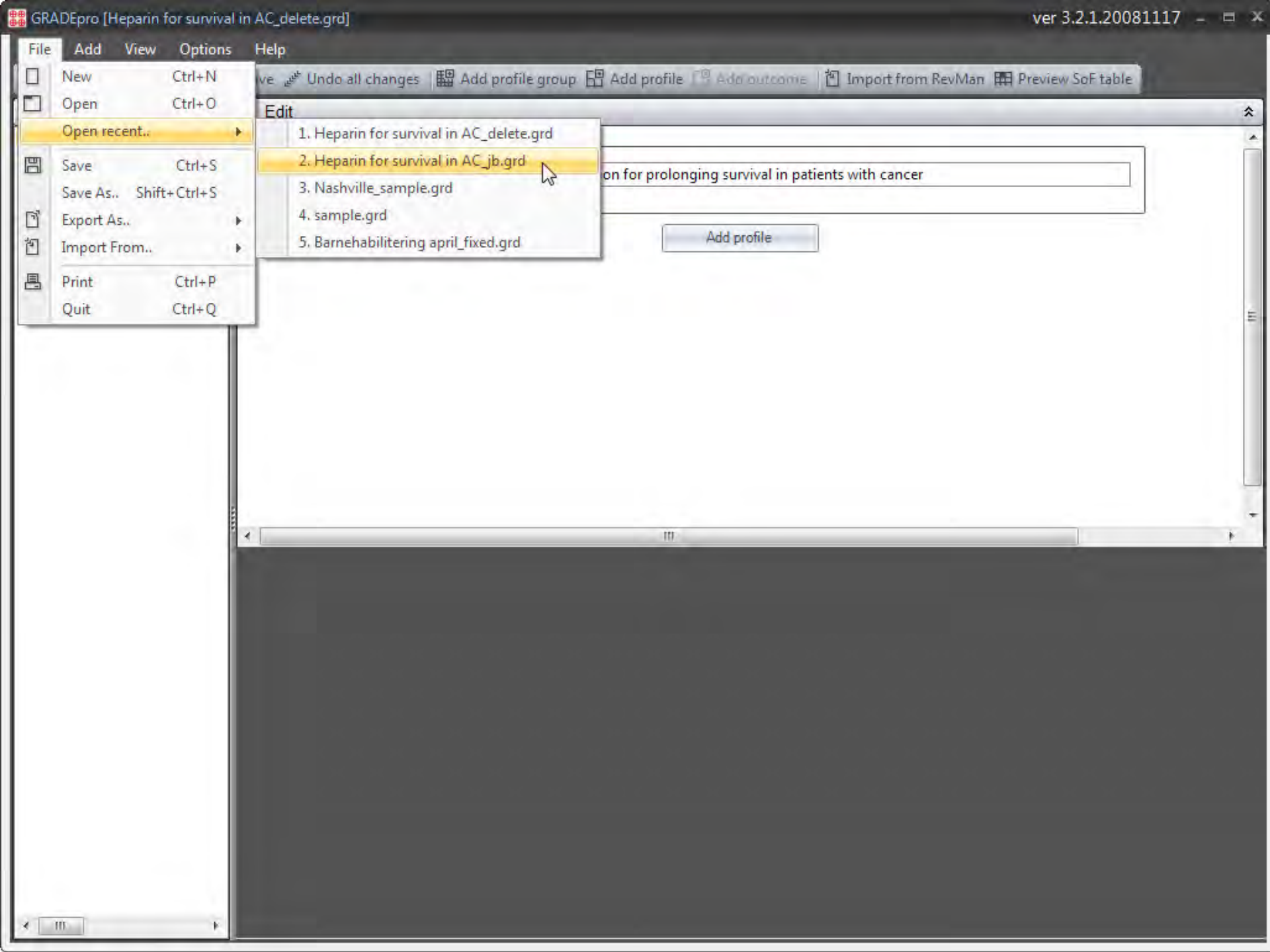
Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Erythropoesis stimulants (epo)	placebo	Relative (95% CI)	Absolute		
All cause mortality (follow-up 4 - 36 months; study follow up (phone calls - death certificates))												
35	randomised trial	no serious limitations	serious[1]	no serious indirectness	serious[2]	none	1008/3825	830/3093 10% 50%	HR 1.11 (1 to 1.22)	25 more per 1000 (from 0 more to 49 more) 10 more per 1,000 36 more per 1,000	⊕⊕OO LOW	CRITICAL
Thromboembolic events												
30	randomised trial	serious[3]	serious[1]	no serious indirectness	no serious imprecision	none	218/3355	112/2737 1% 8%	RR 1.69 (1.36 to 2.1)	28 more per 1000 (from 15 more to 45 more) 6 more per 1,000 55 more per 1,000	⊕⊕OO LOW	CRITICAL
Complete response of tumor to chemotherapy												
5	randomised trial	no serious limitations	serious[1]	no serious indirectness	no serious imprecision	reporting bias[4]	216/344	211/344	RR 1.0 (0.92 to 1.1)	0 fewer per 1000 (from 49 fewer to 61 more)	⊕⊕OO LOW	CRITICAL
Transfusion rates (follow-up 4-26 weeks)												
34	randomised trial	no serious limitations	serious[5]	no serious indirectness	no serious imprecision	none	864/2859	1110/2351 25% 75%	RR 0.63 (0.59 to 0.67)	175 fewer per 1000 (from 156 fewer to 194 fewer) 92 fewer per 1,000 277 fewer per 1,000	⊕⊕⊕O MODERATE	CRITICAL
Increase > 2 mg/dL in Hb (mg/dL) (follow-up 4-20 weeks)												
15	randomised trial	no serious limitations	serious[6]	no serious indirectness	no serious imprecision	strong association[7]	1069/1844	239/1449	RR 3.42 (3.03 to 3.86)	399 more per 1000 (from 335 more to 472 more)	⊕⊕⊕⊕ HIGH	IMPORTANT

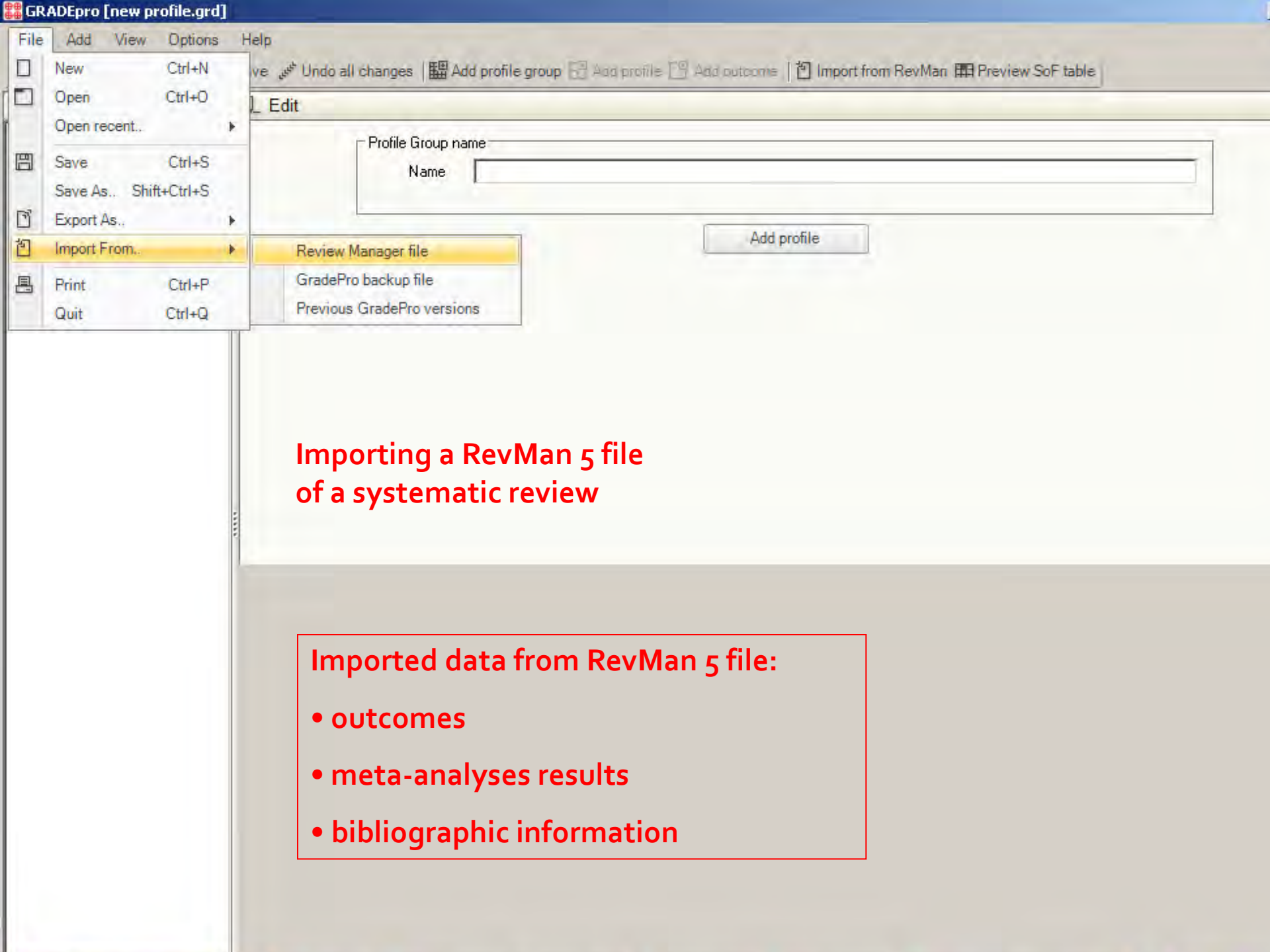
[1] Overall heterogeneity not significant, but underlying clinical heterogeneity due to risk of VTE, treatment regimens, and epo potocols (atarting and stopping Hb).
[2] CI includes no effect and clinically improtant increase in mortality
[3] Criteria for determining and reporting VTE variable in studies; trials reporting varying combinations of DVT, PE, TIA, stroke, and MI
[4] Only 5 trials reported this outcome; does not include the largest trials powered for mortaltiy benefit.
[5]

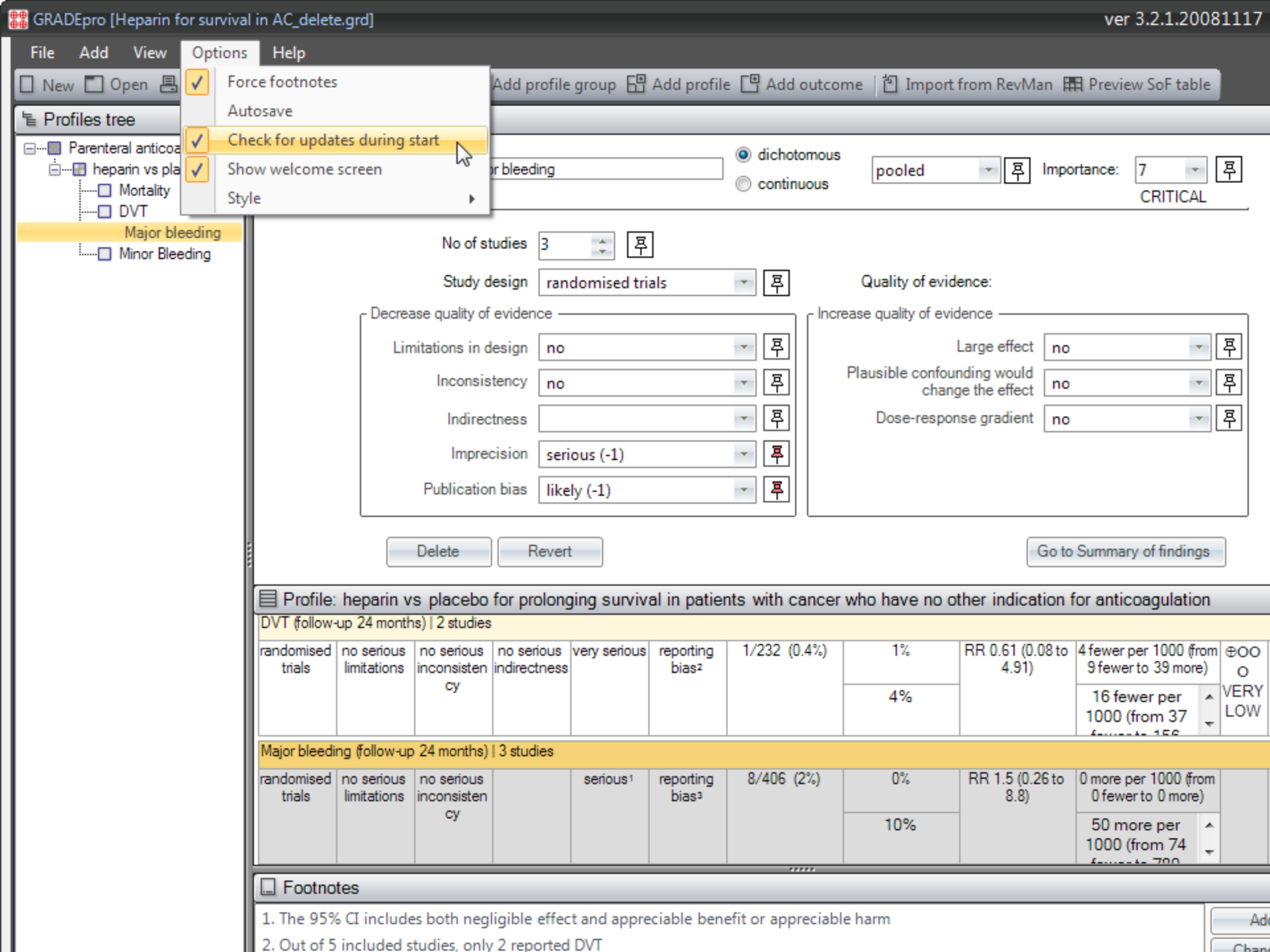
Erythropoesis stimulants (epo) compared to placebo for anemia from cancer chemotherapy

Patient or population: patients with anemia from cancer chemotherapy
Settings: Outpatient cancer treatment
Intervention: Erythropoesis stimulants (epo)
Comparision: placebo

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk placebo	Corresponding risk Erythropoesis stimulants (epo)	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
All cause mortality study follow up (phone calls - death certificates) (follow-up: 4 - 36 months)	**Population**		**HR 1.11** (1 to 1.22)	6918 (35)	⊕⊕OO **low**[1,2]	
	268 per 1000	**293 per 1000** (268 to 317)				
	Low risk population					
	100 per 1000	**110 per 1000** (100 to 121)				
	High risk population					
	500 per 1000	**537 per 1000** (500 to 571)				
Thromboembolic events	**Population**		**RR 1.69** (1.36 to 2.1)	6092 (30)	⊕⊕OO **low**[1,3]	
	41 per 1000	**69 per 1000** (56 to 86)				
	Low risk population					
	10 per 1000	**17 per 1000** (14 to 21)				
	High risk population					
	80 per 1000	**135 per 1000** (109 to 168)				
Complete response of tumor to chemotherapy	**613 per 1000**	**613 per 1000** (564 to 674)	**RR 1.0** (0.92 to 1.1)	688 (5)	⊕⊕OO **low**[1,4]	
Transfusion rates (follow-up: 4-26 weeks)	**Population**		**RR 0.63** (0.59 to 0.67)	5210 (34)	⊕⊕⊕O **moderate**[5]	
	472 per 1000	**297 per 1000** (278 to 316)				
	Low risk population					
	250 per 1000	**158 per 1000** (147 to 168)				


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
New Ctrl+N
Open Ctrl+O
Open recent..
Save Ctrl+S
Save As.. Shift+Ctrl+S
Export As..
Import From..
Print Ctrl+P
Quit Ctrl+Q
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Edit
1. Heparin for survival in AC_delete.grd
2. Heparin for survival in AC_jb.grd
3. Nashville_sample.grd
4. sample.grd
5. Barnehabilitering april_fixed.grd
on for prolonging survival in patients with cancer
Add profile


GRADEpro [new profile.grd]
File
Add
View
Options
Help
New Ctrl+N
Open Ctrl+O
Open recent..
Save Ctrl+S
Save As.. Shift+Ctrl+S
Export As..
Import From..
Print Ctrl+P
Quit Ctrl+Q
Undo all changes
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Edit
Profile Group name
Name
Add profile
Review Manager file
GradePro backup file
Previous GradePro versions
Importing a RevMan 5 file
of a systematic review
Imported data from RevMan 5 file:
• outcomes
• meta-analyses results
• bibliographic information


GRADEpro [Heparin for survival in AC_delete.grd]
ver 3.2.1.20081117
File
Add
View
Options
Help
Force footnotes
Autosave
Check for updates during start
Show welcome screen
Style
New
Open
Add profile group
Add profile
Add outcome
Import from RevMan
Preview SoF table
Profiles tree
Parenteral anticoa
heparin vs pla
Mortality
DVT
Major bleeding
Minor Bleeding
dichotomous
continuous
pooled
Importance:
7
CRITICAL
No of studies
3
Study design
randomised trials
Quality of evidence:
Decrease quality of evidence
Limitations in design
no
Inconsistency
no
Indirectness
Imprecision
serious (-1)
Publication bias
likely (-1)
Increase quality of evidence
Large effect
no
Plausible confounding would change the effect
no
Dose-response gradient
no
Delete
Revert
Go to Summary of findings
Profile: heparin vs placebo for prolonging survival in patients with cancer who have no other indication for anticoagulation
DVT (follow-up 24 months) | 2 studies
randomised trials
no serious limitations
no serious inconsistency
no serious indirectness
very serious
reporting bias2
1/232 (0.4%)
1%
4%
RR 0.61 (0.08 to 4.91)
4 fewer per 1000 (from 9 fewer to 39 more)
16 fewer per 1000 (from 37
⊕OO O VERY LOW
Major bleeding (follow-up 24 months) | 3 studies
randomised trials
no serious limitations
no serious inconsistency
serious1
reporting bias3
8/406 (2%)
0%
10%
RR 1.5 (0.26 to 8.8)
0 more per 1000 (from 0 fewer to 0 more)
50 more per 1000 (from 74
Footnotes
1. The 95% CI includes both negligible effect and appreciable benefit or appreciable harm
2. Out of 5 included studies, only 2 reported DVT

McMaster University
Inspiring Innovation and Discovery

Questions

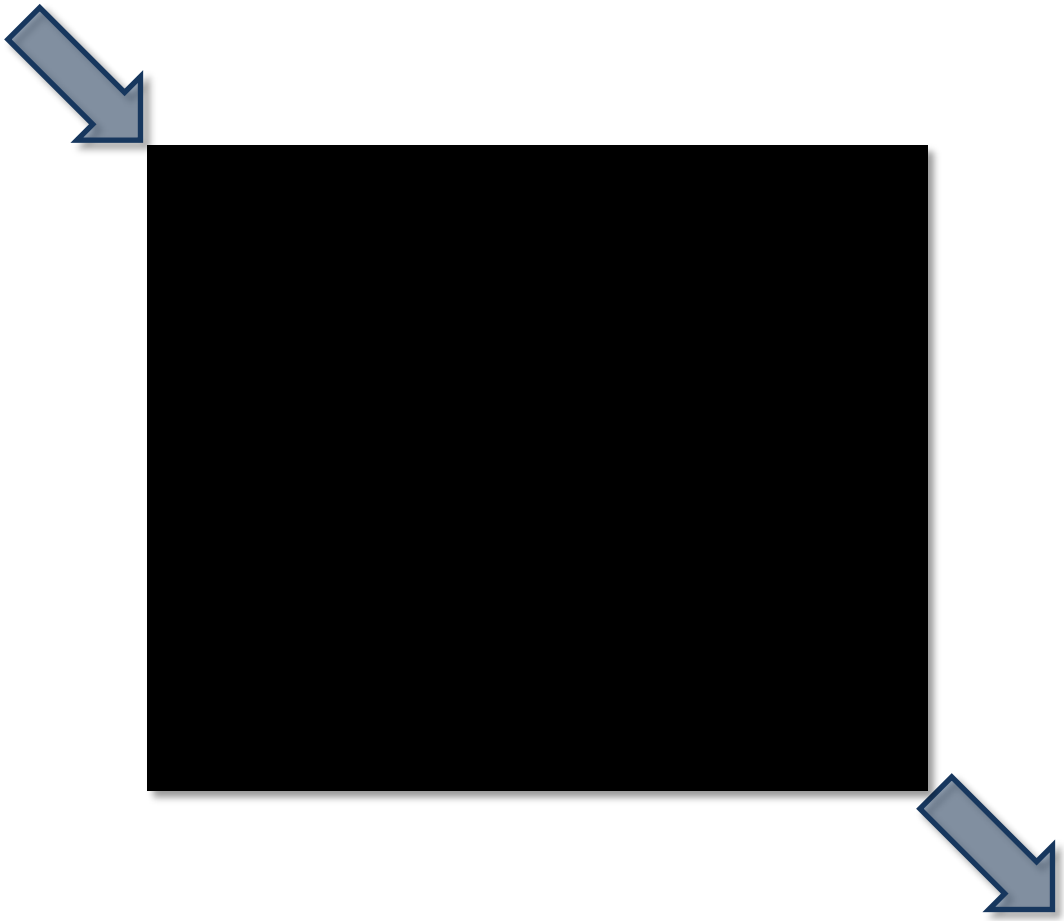
Agenda

McMaster University
Inspiring Innovation and Discovery

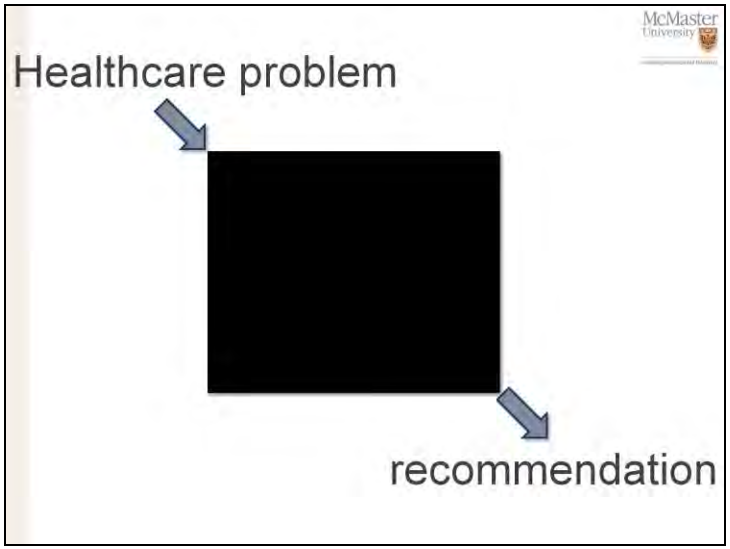
09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

McMaster University
Inspiring Innovation and Discovery

Healthcare problem



recommendation

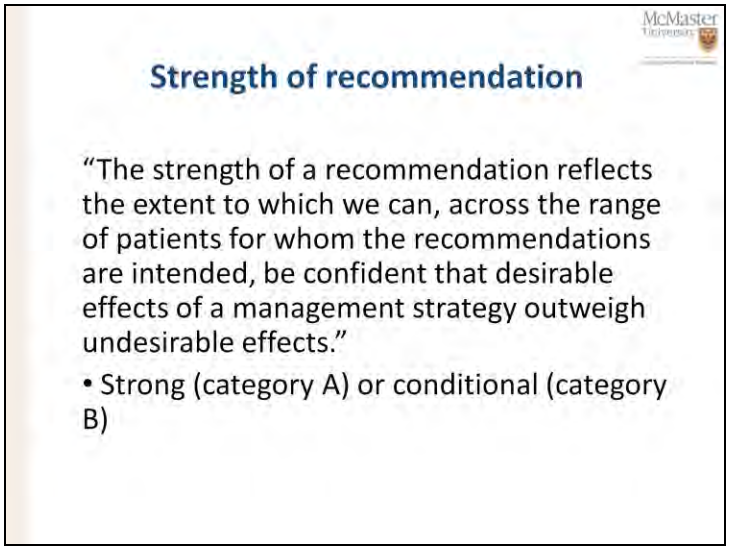


I am now going to speak about how, according to the GRADE approach, one can move from evidence to making recommendations in health care. This truly is a black box in many cases.

Strength of recommendation

"The strength of a recommendation reflects the extent to which we can, across the range of patients for whom the recommendations are intended, be confident that desirable effects of a management strategy outweigh undesirable effects."

- Strong (category A) or conditional (category B)

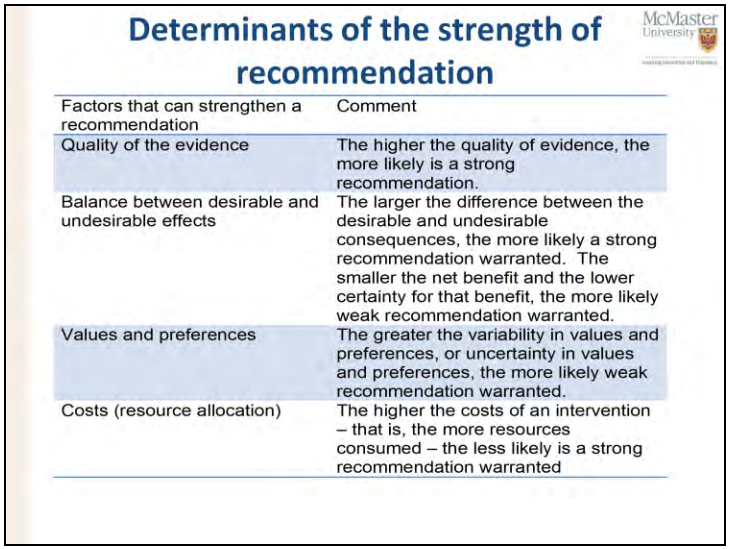


I begin with providing a definition of the strength of recommendation. The strength of recommendation reflects the extent to which we can across a range of patients for whom the recommendations are intended be confident that desirable effects of a management strategy outweigh undesirable effects. Recommendations are made in two categories; for or against an intervention. They can be labeled as strong or conditional. Alternative terms for conditional are weak or discretionary.

Determinants of the strength of recommendation

McMaster University
Inspiring Innovation and Discovery

Factors that can strengthen a recommendation	Comment
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation.
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely weak recommendation warranted.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.
Costs (resource allocation)	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted



The determinants of the strength of recommendation are four, as mentioned previously. The quality of the evidence or the confidence in the estimate of effect that is the higher the quality of evidence the more likely there is a strong recommendation, the balance between benefits and downsides. That is the larger the difference between the benefits and downsides the more likely there is a strong recommendation warranted. The smaller the net benefit and the lower the certainty for that net benefit, the more likely it is that a weak recommendation is warranted. In terms of values and preferences, the greater that variability in values and preferences or the uncertainty in values and preferences or the uncertainty in values and preferences for the outcomes the more likely is the weak recommendation warranted and for resource data and resource utilization the higher the resources required for an intervention, that is the more resources consumed, the less likely is a strong recommendation warranted, in particular if there is a small net benefit.

Trends in guideline production

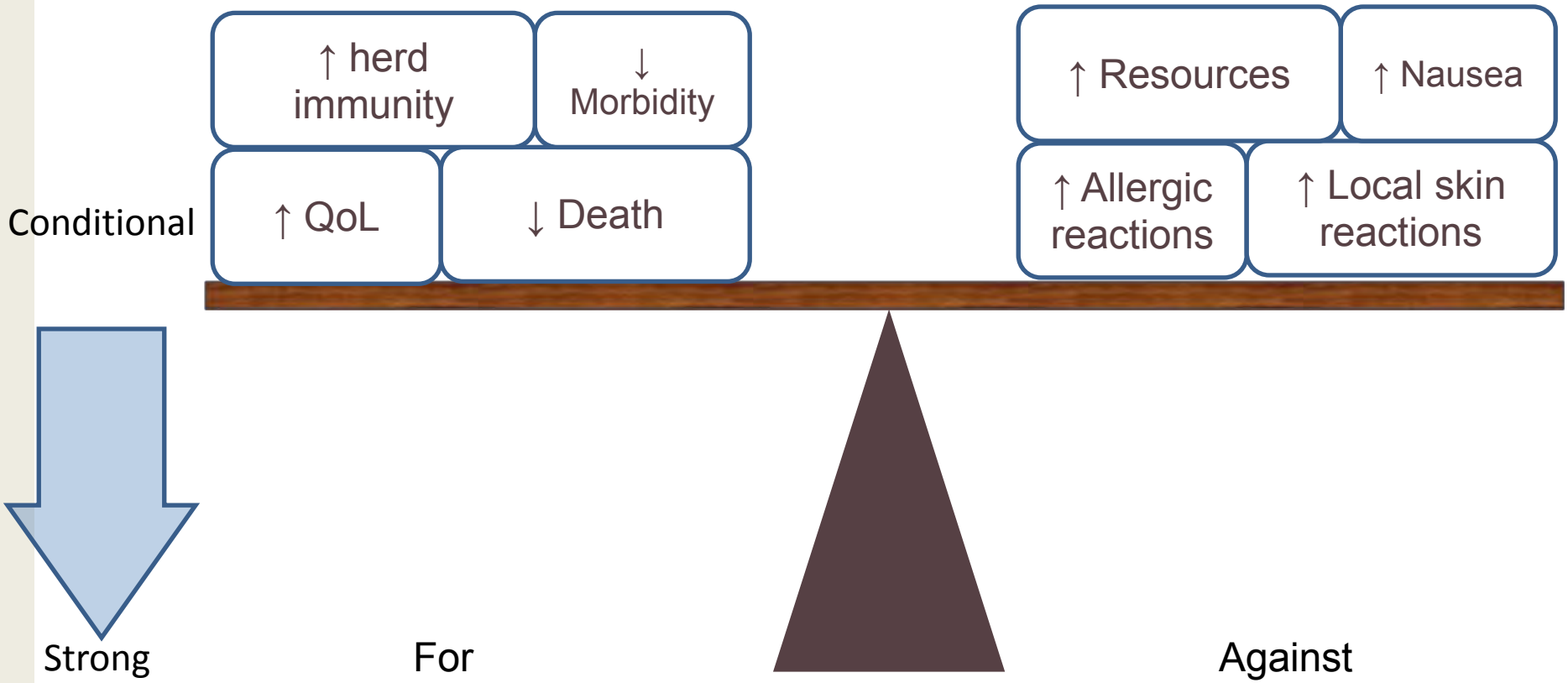
(AHA guidelines, Tricoci JAMA 2009)

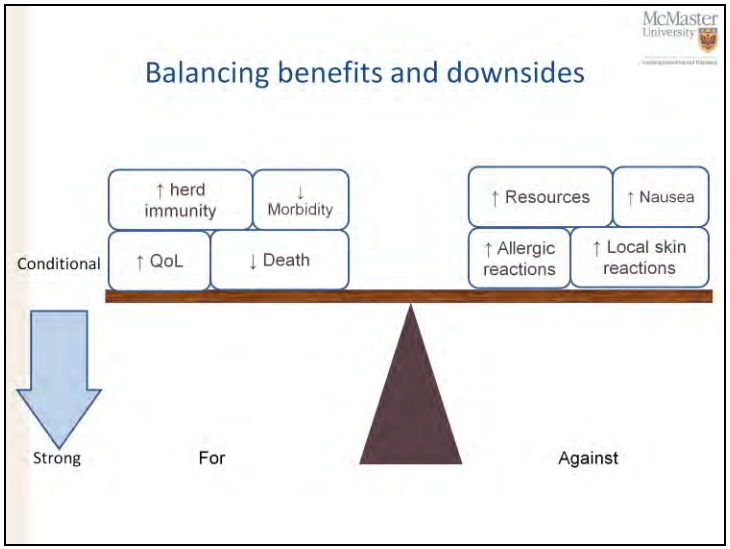
- Recommendations are increasing in size with every update (+48% form first version)
- Levels (quality of evidence: only a minority of recommendations are based in good evidence (11%) and half (48%) on low quality
- Recommendations with level of evidence A are mostly concentrated in class I (strong recommendation or useful and effective), but only 245 of 1305 class I recommendations have level of evidence A (median, 19%)

McMaster University
Inspiring Innovation and Discovery

How to improve transparency in going from evidence to recomendations

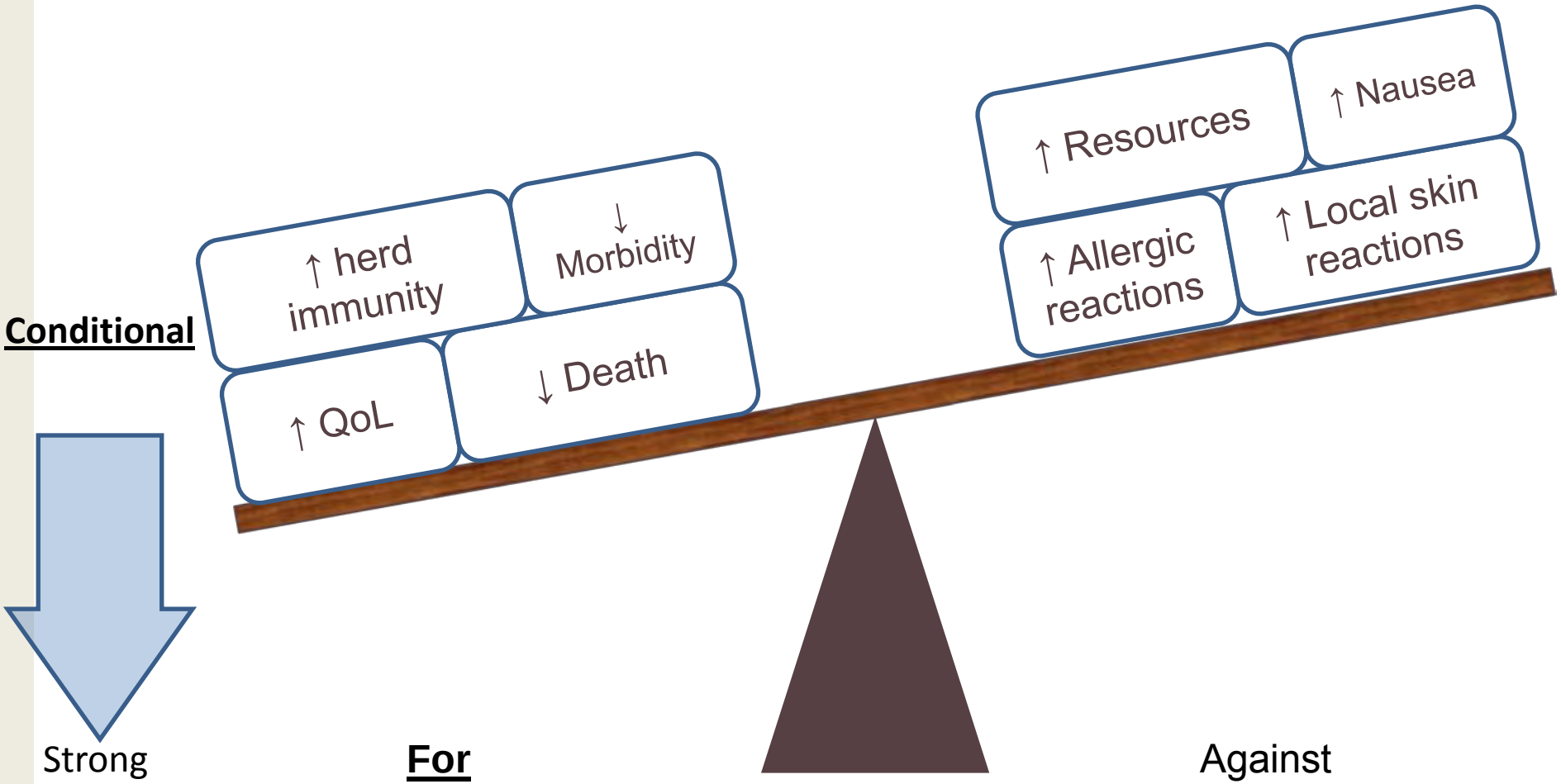
Balancing benefits and downsides

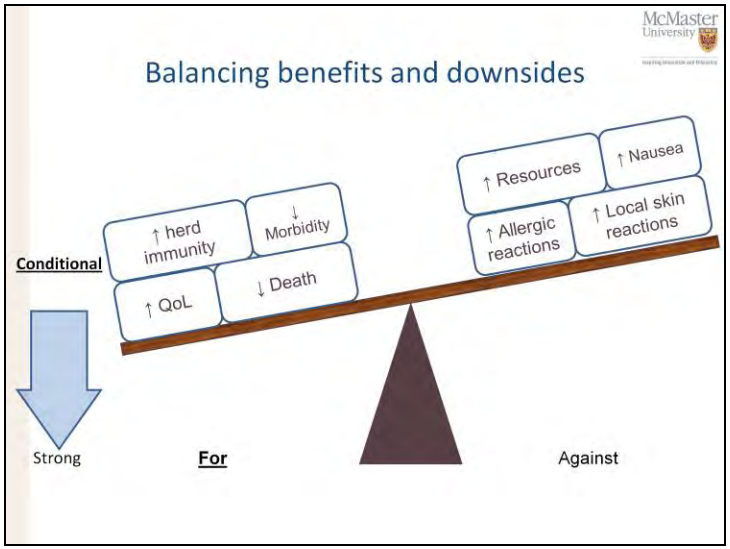




So this can be conceptualized as balancing the benefits and the down sides, where the benefits obviously include a value judgment that is how important the outcome is. On this balance, therefore, each square represents a combination of the magnitude of the effect and the importance of that effect. This balance then can be evaluated either through an informed judgment or more or less complicated decision analysis. The quality of the evidence is considered by assigning an overall quality of the evidence. That is when the quality of evidence is high we have a lot of uncertainty in the balance that is evaluated here when the quality of evidence is low or very low, we have much less certainty about how this balance would behave in the real world. Than according to how this balance behaves, we offer recommendations.

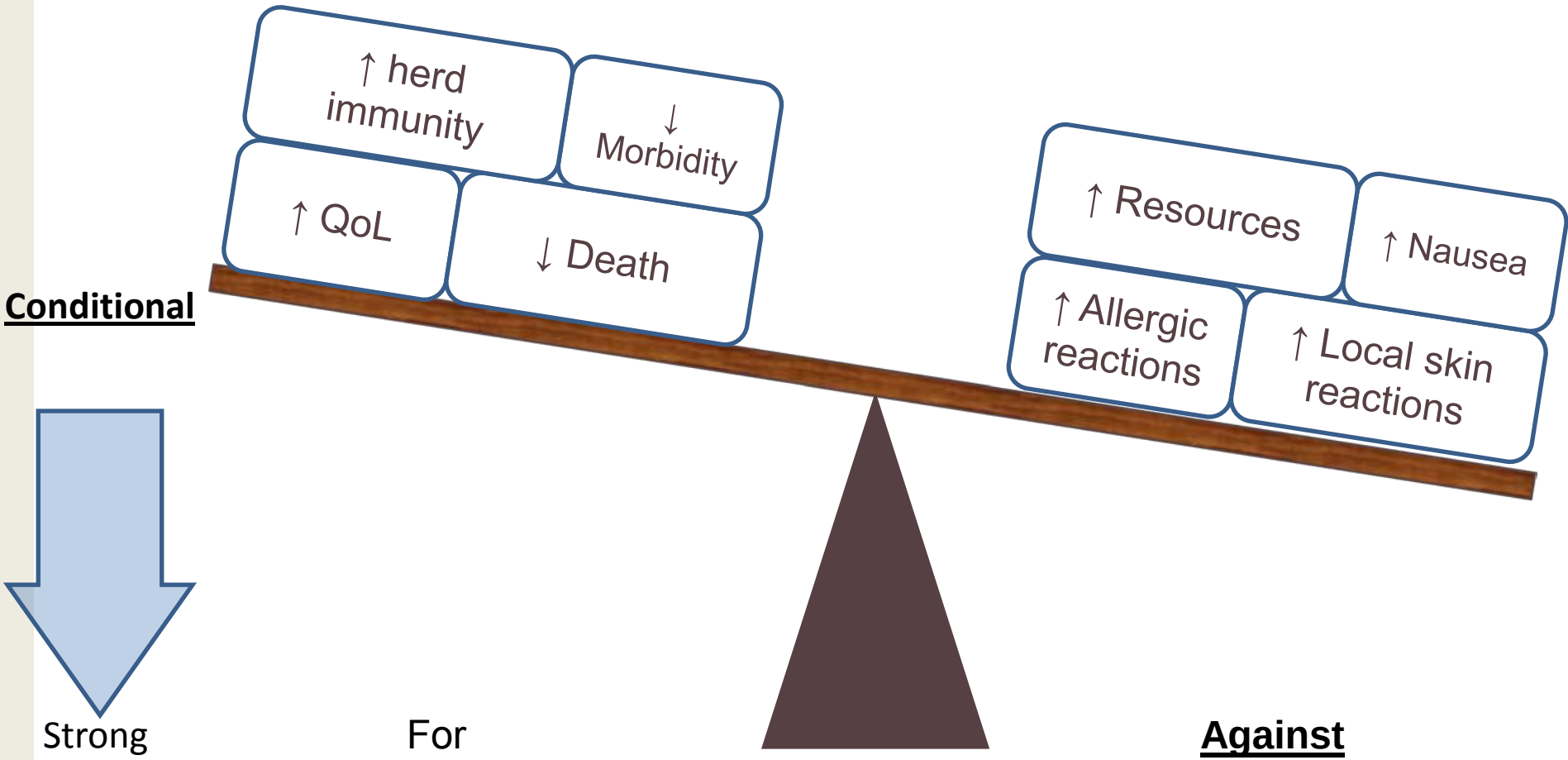
Balancing benefits and downsides

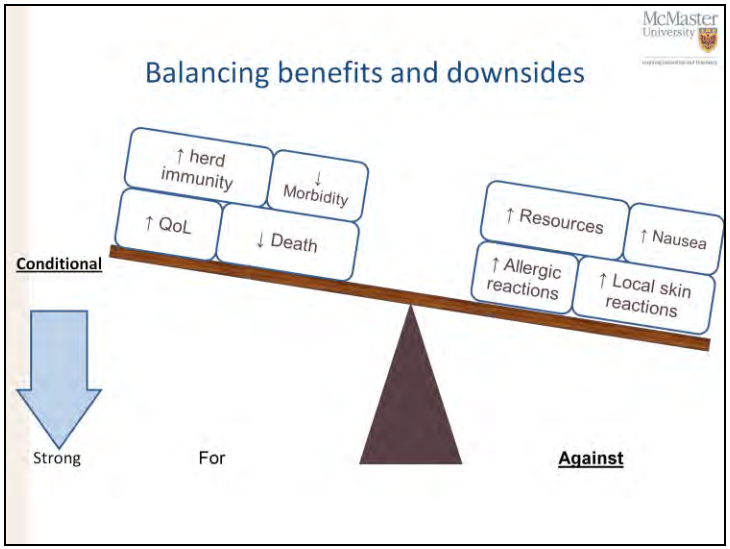




If the benefits slightly outweigh the downsides we make a condition recommendation for an intervention.

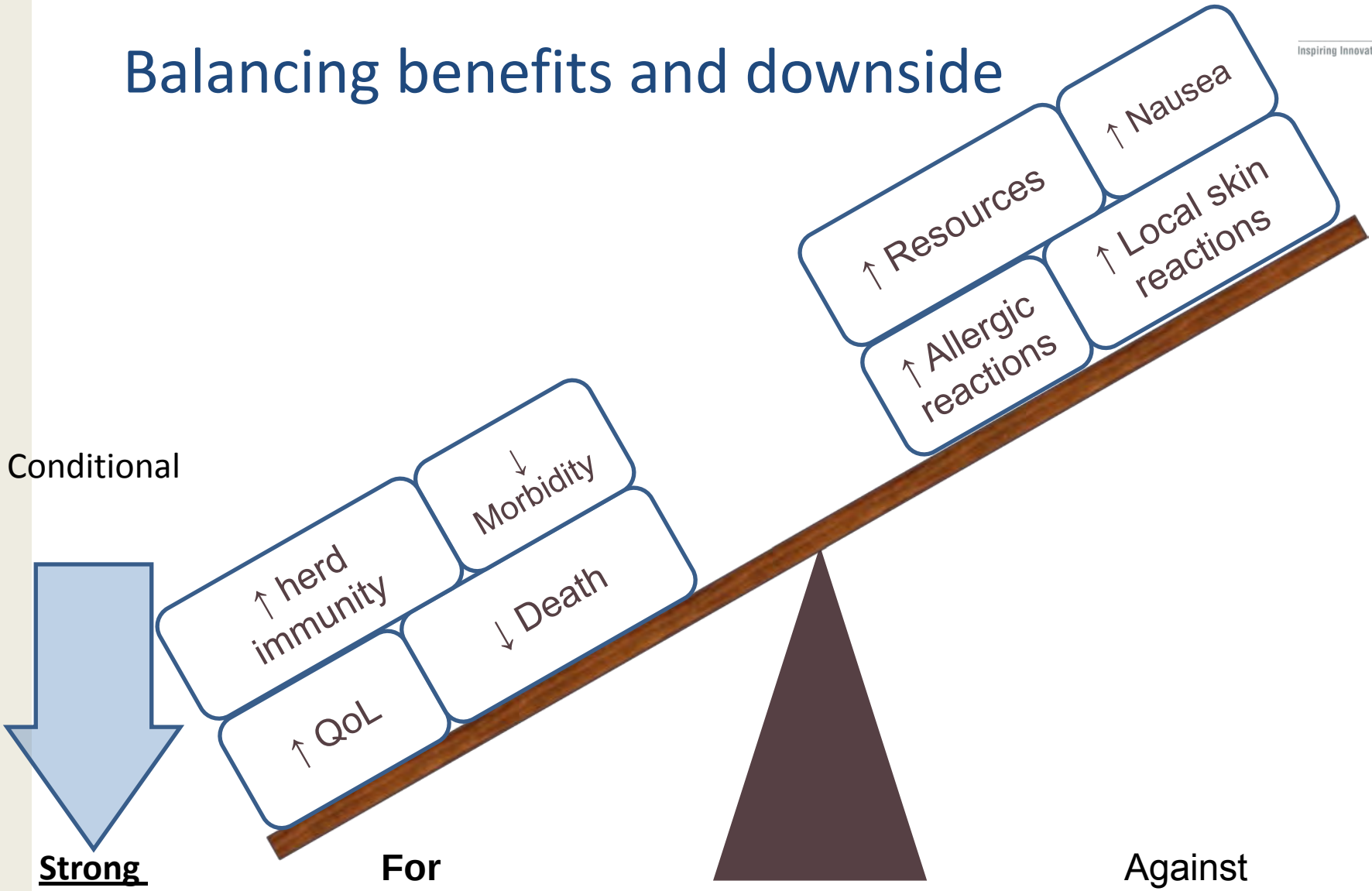
Balancing benefits and downsides





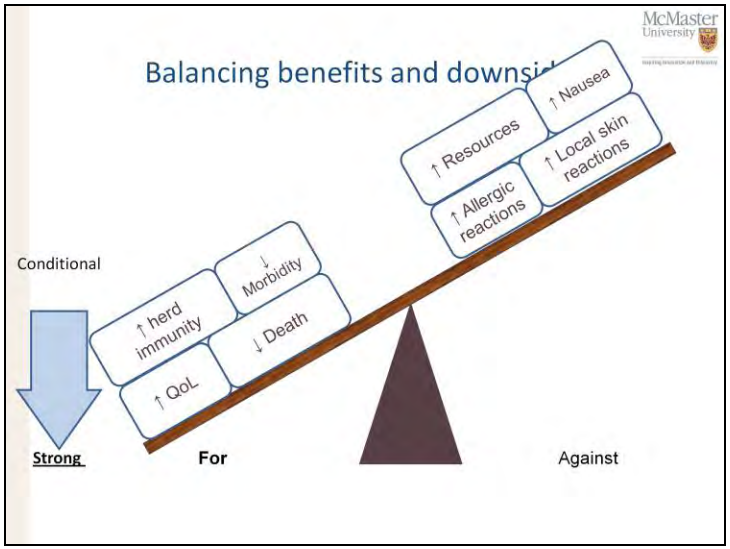
If the benefits slightly outweigh the benefits we make a conditional recommendation against an intervention.

Balancing benefits and downside



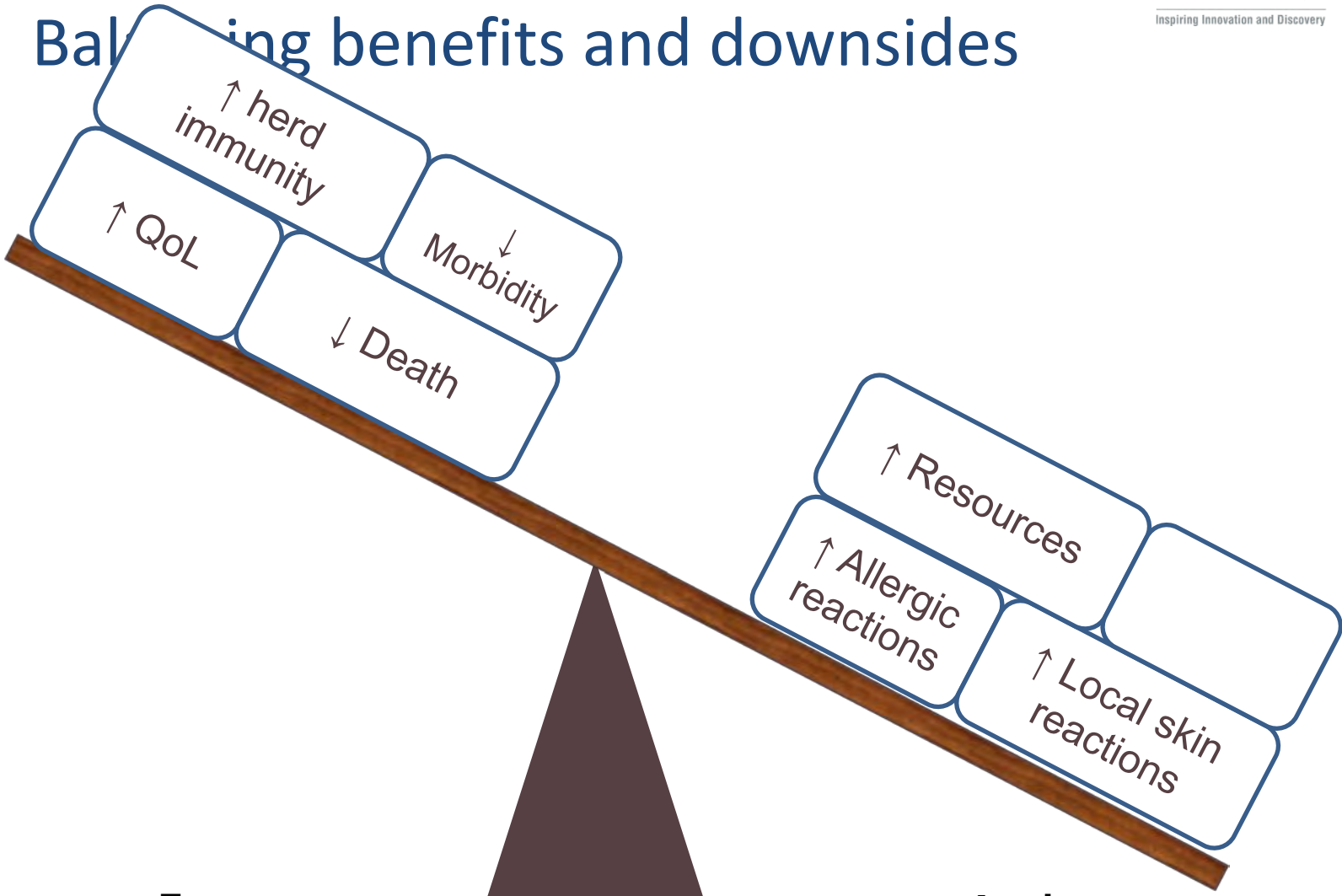
Conditional

Strong **For** Against



If the balance is clearly in favor of the benefits we make a strong recommendation for an intervention

Bal ing benefits and downsides

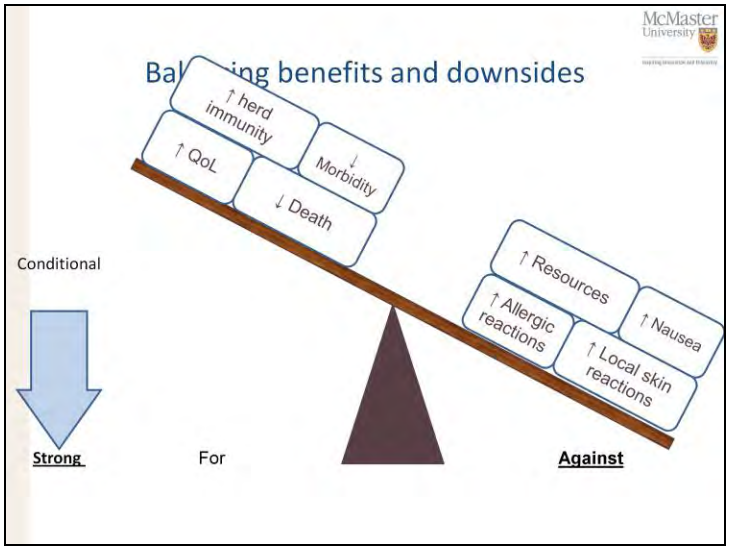


Conditional

Strong

For

Against



and if the downsides clearly outweigh the benefits we make a strong recommendation against an intervention. Please remember that diagnostic tests and strategies are considered interventions in the large context of GRADE.

McMaster University
Inspiring Innovation and Discovery

Examples of recommendations using GRADE

Examples of transparency

Case scenario

A 13 year old girl who lives in rural Indonesia presented with flu symptoms and developed severe respiratory distress over the course of the last 2 days. She required intubation. The history reveals that she shares her living quarters with her parents and her three siblings. At night the family's chicken stock shares this room too and several chicken had died unexpectedly a few days before the girl fell sick.

Methods – WHO Rapid Advice Guidelines for Avian Flu

- Applied findings of a recent systematic evaluation of guideline development for WHO/ACHR
- Group composition (including panel of 13 voting members):
  - clinicians who treated influenza A(H5N1) patients
  - infectious disease experts
  - basic scientists
  - public health officers
  - methodologists
- Independent scientific reviewers:
  - Identified systematic reviews, recent RCTs, case series, animal studies related to H5N1 infection

Oseltamivir for Avian Flu

McMaster University
Inspiring Innovation and Discovery

Summary of findings:

- No clinical trial of oseltamivir for treatment of H5N1 patients.
- 4 systematic reviews and health technology assessments (HTA) reporting on 5 studies of oseltamivir in seasonal influenza.
  - Hospitalization: OR 0.22 (0.02 – 2.16)
  - Pneumonia: OR 0.15 (0.03 - 0.69)
- 3 published case series.
- Many in vitro and animal studies.
- No alternative that was more promising at present.
- Cost: 40$ per treatment course

From evidence to recommendation

McMaster University
Inspiring Innovation and Discovery

Factors that can strengthen a recommendation	Comment
Quality of the evidence	Very low quality evidence
Balance between desirable and undesirable effects	Uncertain, but small reduction in relative risk still leads to large absolute effect
Values and preferences	Little variability and clear
Costs (resource allocation)	Low cost under non-pandemic conditions

Example: Oseltamivir for Avian Flu

McMaster University
Inspiring Innovation and Discovery

Recommendation: In patients with confirmed or strongly suspected infection with avian influenza A (H5N1) virus, clinicians should administer oseltamivir treatment as soon as possible (strong recommendation, very low quality evidence).

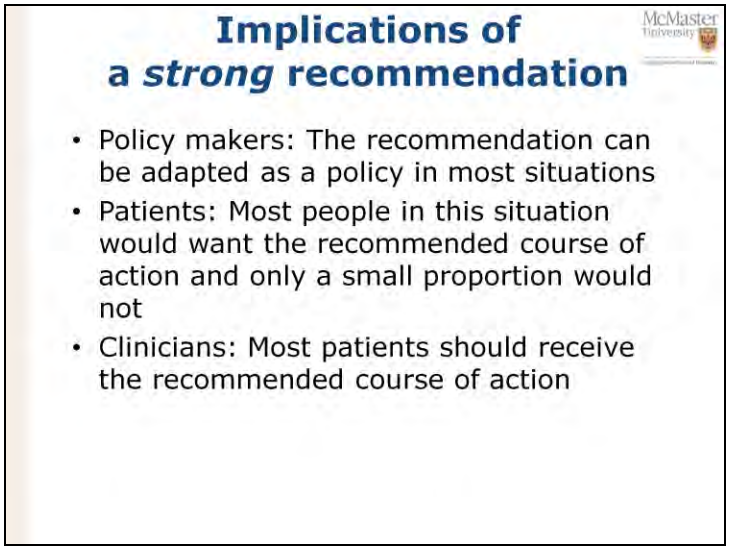
Remarks: This recommendation places a high value on the prevention of death in an illness with a high case fatality. It places relatively low values on adverse reactions, the development of resistance and costs of treatment.

Schunemann et al. The Lancet ID, 2007

Implications of a *strong* recommendation

McMaster University
Inspiring Innovation and Discovery

- Policy makers: The recommendation can be adapted as a policy in most situations
- Patients: Most people in this situation would want the recommended course of action and only a small proportion would not
- Clinicians: Most patients should receive the recommended course of action

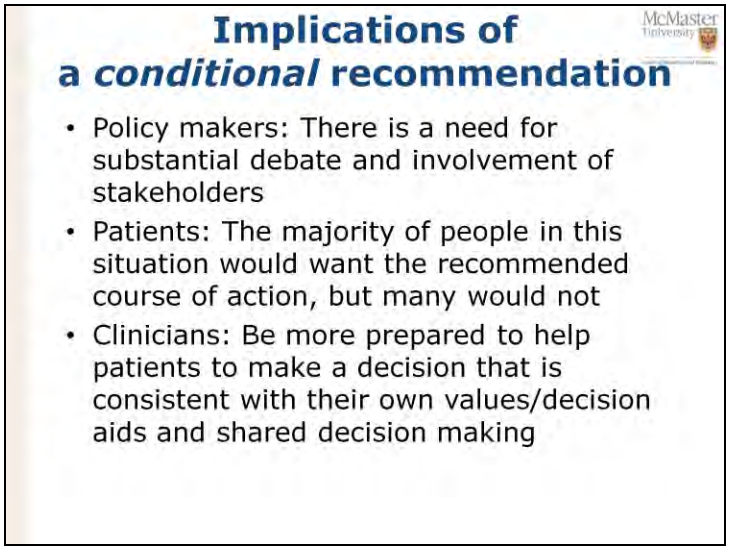


The implications of a strong recommendation are for patients that most people in this situation would want the recommended course of action and only a small proportion would not. For clinicians or health care providers it means that most patients should receive the recommended course of action for policy makers or those advising quality indicators the recommendation could be adapted as a policy in most situations.

Implications of a *conditional* recommendation

McMaster University
Inspiring Innovation and Discovery

- Policy makers: There is a need for substantial debate and involvement of stakeholders
- Patients: The majority of people in this situation would want the recommended course of action, but many would not
- Clinicians: Be more prepared to help patients to make a decision that is consistent with their own values/decision aids and shared decision making

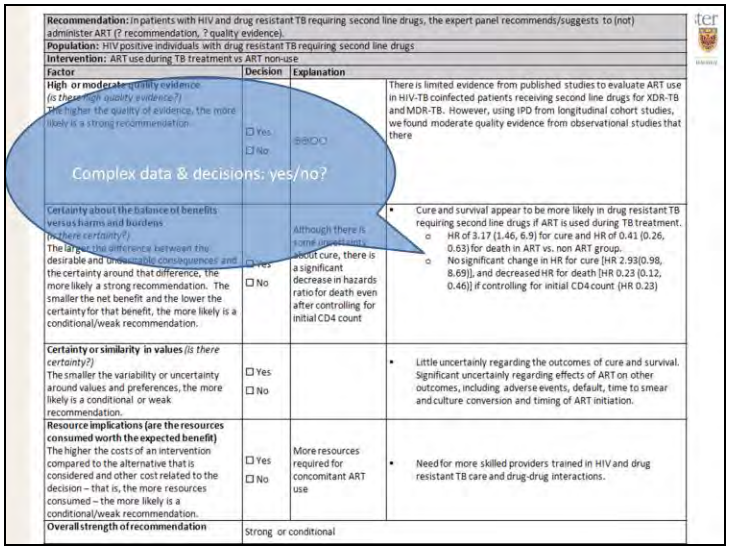


The implications of a weaker conditional recommendation are that patients in the majority would, if they were confronted with the situation, want the recommended course of action but many would not want the recommended course of actions. For clinicians or health care providers it means that they should be more prepared to help patients or the target population to make a decision that is consistent with their own values. Decision aids and shared decision making are very appropriate under those circumstances or even more appropriate, and for policy makers or those devising quality indicators it means that there is a need for substantial debate and involvement of stakeholders. It also means that as a quality indicator a weak recommendation would only serve if the quality indicator was that an informed decision has been made or that a decision aid for instance was used.

Recommendation: In patients with HIV and drug resistant TB requiring second line drugs, the expert panel recommends/suggests to (not) administer ART (? recommendation, ? quality evidence).			
Population: HIV positive individuals with drug resistant TB requiring second line drugs			
Intervention: ART use during TB treatment vs ART non-use			
Factor	**Decision**	**Explanation**	
High or moderate quality evidence *(is there high quality evidence?)* The higher the quality of evidence, the more likely is a strong recommendation.	☐ Yes ☐ No	⊕⊕◯◯	There is limited evidence from published studies to evaluate ART use in HIV-TB coinfected patients receiving second line drugs for XDR-TB and MDR-TB. However, using IPD from longitudinal cohort studies, we found moderate quality evidence from observational studies that there
Certainty about the balance of benefits versus harms and burdens *(is there certainty?)* The larger the difference between the desirable and undesirable consequences and the certainty around that difference, the more likely a strong recommendation. The smaller the net benefit and the lower the certainty for that benefit, the more likely is a conditional/weak recommendation.	☐ Yes ☐ No	Although there is some uncertainty about cure, there is a significant decrease in hazards ratio for death even after controlling for initial CD4 count	▪ Cure and survival appear to be more likely in drug resistant TB requiring second line drugs if ART is used during TB treatment. o HR of 3.17 (1.46, 6.9) for cure and HR of 0.41 (0.26, 0.63) for death in ART vs. non ART group. o No significant change in HR for cure [HR 2.93(0.98, 8.69)], and decreased HR for death [HR 0.23 (0.12, 0.46)] if controlling for initial CD4 count (HR 0.23)
Certainty or similarity in values *(is there certainty?)* The smaller the variability or uncertainty around values and preferences, the more likely is a conditional or weak recommendation.	☐ Yes ☐ No		▪ Little uncertainly regarding the outcomes of cure and survival. Significant uncertainly regarding effects of ART on other outcomes, including adverse events, default, time to smear and culture conversion and timing of ART initiation.
Resource implications (are the resources consumed worth the expected benefit) The higher the costs of an intervention compared to the alternative that is considered and other cost related to the decision – that is, the more resources consumed – the more likely is a conditional/weak recommendation.	☐ Yes ☐ No	More resources required for concomitant ART use	▪ Need for more skilled providers trained in HIV and drug resistant TB care and drug-drug interactions.
Overall strength of recommendation	Strong or conditional		

Complex data & decisions: yes/no?

McMaster University
Inspiring Innovation and Discovery



Recommendation: In patients with HIV and drug resistant TB requiring second line drugs, the expert panel recommends/suggests to (not) administer ART (? recommendation, ? quality evidence).			
Population: HIV positive individuals with drug resistant TB requiring second line drugs			
Intervention: ART use during TB treatment vs ART non-use			
Factor	**Decision**	**Explanation**	
High or moderate quality evidence *(is there high quality evidence?)* The higher the quality of evidence, the more likely is a strong recommendation	☐ Yes ☐ No		There is limited evidence from published studies to evaluate ART use in HIV-TB coinfected patients receiving second line drugs for XDR-TB and MDR-TB. However, using IPD from longitudinal cohort studies, we found moderate quality evidence from observational studies that there
Certainty about the balance of benefits versus harms and burdens *(is there certainty?)* The larger the difference between the desirable and undesirable consequences and the certainty around that difference, the more likely a strong recommendation. The smaller the net benefit and the lower the certainty for that benefit, the more likely is a conditional/weak recommendation.	☐ Yes ☐ No	Although there is some uncertainty about cure, there is a significant decrease in hazards ratio for death even after controlling for initial CD4 count	• Cure and survival appear to be more likely in drug resistant TB requiring second line drugs if ART is used during TB treatment. ○ HR of 3.17 (1.46, 6.9) for cure and HR of 0.41 (0.26, 0.63) for death in ART vs. non ART group. ○ No significant change in HR for cure [HR 2.93(0.98, 8.69)], and decreased HR for death [HR 0.23 (0.12, 0.46)] *if* controlling for initial CD4 count (HR 0.23)
Certainty or similarity in values *(is there certainty?)* The smaller the variability or uncertainty around values and preferences, the more likely is a conditional or weak recommendation.	☐ Yes ☐ No		• Little uncertainly regarding the outcomes of cure and survival. Significant uncertainly regarding effects of ART on other outcomes, including adverse events, default, time to smear and culture conversion and timing of ART initiation.
Resource implications (are the resources consumed worth the expected benefit) The higher the costs of an intervention compared to the alternative that is considered and other cost related to the decision – that is, the more resources consumed – the more likely is a conditional/weak recommendation.	☐ Yes ☐ No	More resources required for concomitant ART use	• Need for more skilled providers trained in HIV and drug resistant TB care and drug-drug interactions.
Overall strength of recommendation	Strong or conditional		

Complex data & decisions: yes/no?

Various organizations have started to use this type of evidence to recommendation or decision tables; this is an example from a WHO guideline that deals with treating patients with tuberculosis. The four factors are evaluated and listed in the left hand column. In the right hand column there is information in regards to how the evidence addresses this particular category. The explanation then provides a brief summary of the guideline panels judgment and decision and the yes and no decision refers to whether there, for instance, is high or moderate quality evidence or whether there is certainty about the benefits and downsides. At the end on overall recommendation is made, the strength of which is determined by whether the panel has a great deal of certainty or whether the quality of evidence is high. Under those circumstances when there are many yes answers a strong recommendation is more likely.

Recommendation

- The Guidelines Group recommends that fluoroquinolones are / not used in the treatment of all patients with MDR

  (Strong(conditional) recommendation/ low(moderate, high) grade of evidence)

Issues in guideline development for immunization

McMaster University
Inspiring Innovation and Discovery

- Causation versus effects of intervention
  - Causation not equivalent to efficacy of interventions
  - Bradford Hill
    - Nearly half a century old – tablet from the mountain?
- Harms caused by interventions
  - Assumption is that removal of vaccine (or no exposure) leads to NO adverse effects
- How confident can one be that removal of the exposure is effective in preventing disease?
  - Whether immunization or environmental factors: will depend on the intervention to remove exposure

McMaster University
Inspiring Innovation and Discovery

Current state of recommendations

INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 78 (2009) 354–363

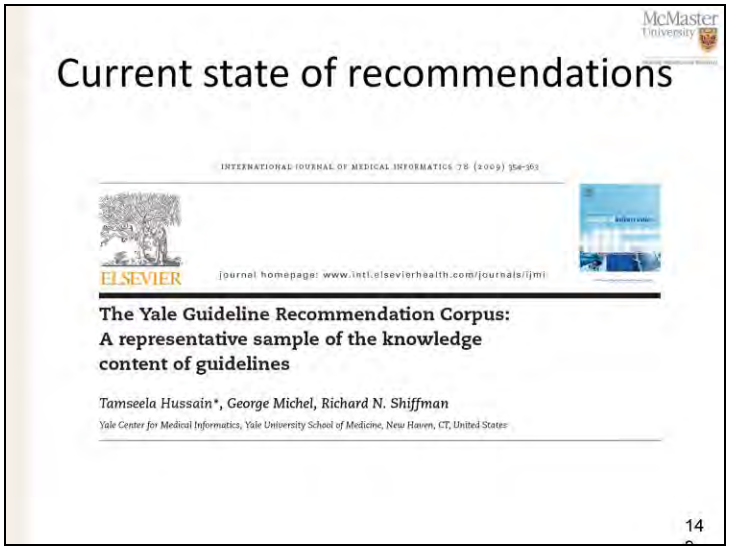
ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/ijmi

The Yale Guideline Recommendation Corpus: A representative sample of the knowledge content of guidelines

Tamseela Hussain, George Michel, Richard N. Shiffman*

Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, United States



This is an interesting piece of work describing what is being done in this field in terms of describing recommendations.

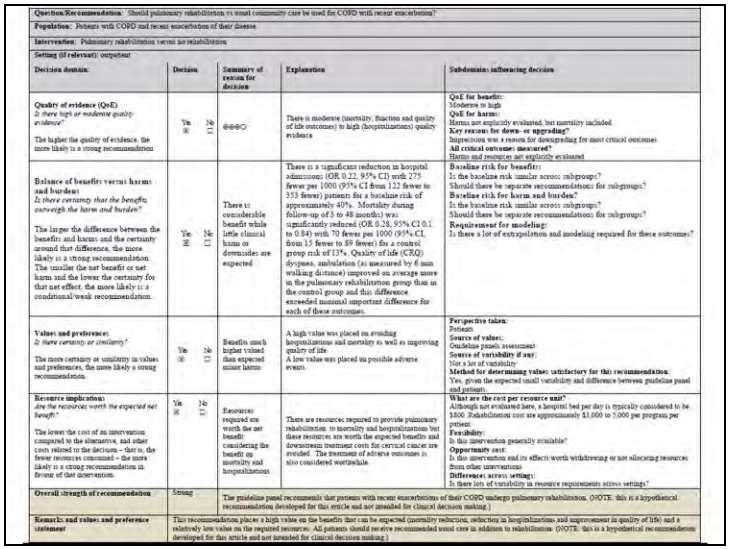
Current state of recommendations

- Reviewed 7527 recommendations
  - 1275 randomly selected
- Inconsistency across/within
- 31.6% did not recommendations clearly
  - Most of them not written as executable actions
- 52.7% did not indicated strength

Recommendation

- The Guideline Group recommends rapid DST testing for resistance to INH and RIF or RIF alone over conventional testing or no testing at the time of diagnosis of TB (conditional, ⊕⊕◯◯ /low quality evidence).
- Values and preferences: A high value was placed on outcomes such as preventing death and transmission of MDR as a result of delayed diagnosis as well as avoiding spending resources.

Question/Recommendation: Should pulmonary rehabilitation vs usual community care be used for COPD with recent exacerbation?				
Population: Patients with COPD and recent exacerbation of their disease				
Intervention: Pulmonary rehabilitation versus no rehabilitation				
Setting (if relevant): outpatient				
Decision domain:	**Decision**	**Summary of reason for decision**	**Explanation**	**Subdomains influencing decision**
Quality of evidence (QoE) *Is there high or moderate quality evidence?* The higher the quality of evidence, the more likely is a strong recommendation	Yes ☒ No ☐	⊕⊕⊕O	There is moderate (mortality, function and quality of life outcomes) to high (hospitalizations) quality evidence	**QoE for benefits:** Moderate to high **QoE for harms:** Harms not explicitly evaluated, but mortality included **Key reasons for down- or upgrading?** Imprecision was a reason for downgrading for most critical outcomes **All critical outcomes measured?** Harms and resources not explicitly evaluated
Balance of benefits versus harms and burdens *Is there certainty that the benefits outweigh the harm and burden?* The larger the difference between the benefits and harms and the certainty around that difference, the more likely is a strong recommendation. The smaller the net benefit or net harm and the lower the certainty for that net effect, the more likely is a conditional/weak recommendation.	Yes ☒ No ☐	There is considerable benefit while little clinical harm or downsides are expected	There is a significant reduction in hospital admissions (OR 0.22, 95% CI) with 275 fewer per 1000 (95% CI from 122 fewer to 353 fewer) patients for a baseline risk of approximately 40%. Mortality during follow-up of 3 to 48 months) was significantly reduced (OR 0.28, 95% CI 0.1 to 0.84) with 70 fewer per 1000 (95% CI, from 15 fewer to 89 fewer) for a control group risk of 13%. Quality of life (CRQ) dyspnea, ambulation (as measured by 6 min walking distance) improved on average more in the pulmonary rehabilitation group than in the control group and this difference exceeded minimal important difference for each of these outcomes.	**Baseline risk for benefits:** Is the baseline risk similar across subgroups? Should there be separate recommendations for subgroups? **Baseline risk for harm and burden?** Is the baseline risk similar across subgroups? Should there be separate recommendations for subgroups? **Requirement for modeling:** Is there a lot of extrapolation and modeling required for these outcomes?
Values and preferences *Is there certainty or similarity?* The more certainty or similarity in values and preferences, the more likely a strong recommendation.	Yes ☒ No ☐	Benefits much higher valued than expected minor harms.	A high value was placed on avoiding hospitalizations and mortality as well as improving quality of life. A low value was placed on possible adverse events.	**Perspective taken:** Patients **Source of values:** Guideline panels assessment **Source of variability if any:** Not a lot of variability **Method for determining values satisfactory for this recommendation:** Yes, given the expected small variability and difference between guideline panel and patients.
Resource implications *Are the resources worth the expected net benefit?* The lower the cost of an intervention compared to the alternative, and other costs related to the decision – that is, the fewer resources consumed – the more likely is a strong recommendation in favour of that intervention.	Yes ☒ No ☐	Resources required are worth the net benefit considering the benefit on mortality and hospitalizations.	There are resources required to provide pulmonary rehabilitation to mortality and hospitalitzations but these resources are worth the expected benefits and downstream treatment costs for cervical cancer are avoided. The treatment of adverse outcomes is also considered worthwhile.	**What are the cost per resource unit?** Although not evaluated here, a hospital bed per day is typically considered to be $800. Rehabilitation cost are approximately $3,000 to 5,000 per program per patient. **Feasibility:** Is this intervention generally available? **Opportunity cost:** Is this intervention and its effects worth withdrawing or not allocating resources from other interventions **Differences across settings:** Is there lots of variability in resource requirements across settings?
Overall strength of recommendation	Strong	The guideline panel recommends that patients with recent exacerbations of their COPD undergo pulmonary rehabilitation. (NOTE: this is a hypothetical recommendation developed for this article and not intended for clinical decision making.)		
Remarks and values and preference statement	This recommendation places a high value on the benefits that can be expected (mortality reduction, reduction in hospitalizations and improvement in quality of life) and a relatively low value on the required resources. All patients should receive recommended usual care in addition to rehabilitation. (NOTE: this is a hypothetical recommendation developed for this article and not intended for clinical decision making.)			



Question/Recommendation: Should pulmonary rehabilitation vs usual community care be used for COPD with recent exacerbation?				
Population: Patients with COPD and recent exacerbation of their disease				
Intervention: Pulmonary rehabilitation versus no rehabilitation				
Setting (if relevant): outpatient				
Decision domain:	**Decision**	**Summary of reason for decision**	**Explanation**	**Subdomains influencing decision**
Quality of evidence (QoE) *Is there high or moderate quality evidence?* The higher the quality of evidence, the more likely is a strong recommendation	Yes ☒ No ☐	⊕⊕⊕◯	There is moderate (mortality, fraction and quality of life outcomes) to high (hospitalizations) quality evidence	**QoE for benefit:** Moderate to high **QoE for harm:** Harms not explicitly evaluated, but mortality included **Key reason for down- or upgrading?** Imprecision was a reason for downgrading for most critical outcomes **All critical outcomes measured?** Harms and resources not explicitly evaluated
Balance of benefits versus harms and burdens *Is there certainty that the benefits outweigh the harm and burden?* The larger the difference between the benefits and harms and the certainty around that difference, the more likely is a strong recommendation. The smaller the net benefit or net harm and the lower the certainty for that net effect, the more likely is a conditional/weak recommendation.	Yes ☒ No ☐	There is considerable benefit while little clinical harm or downsides are expected	There is a significant reduction in hospital admissions (OR 0.22, 95% CI) with 275 fewer per 1000 (95% CI from 122 fewer to 353 fewer) patients for a baseline risk of approximately 40%. Mortality during follow-up of 3 to 48 months) was significantly reduced (OR 0.28, 95% CI 0.1 to 0.84) with 70 fewer per 1000 (95% CI, from 15 fewer to 89 fewer) for a control group risk of 13%. Quality of life (CRQ) dyspnea, ambulation (as measured by 6 min walking distance) improved on average more in the pulmonary rehabilitation group than in the control group and this difference exceeded minimal important difference for each of these outcomes.	**Baseline risk for benefit:** Is the baseline risk similar across subgroups? Should there be separate recommendations for subgroups? **Baseline risk for harm and burden:** Is the baseline risk similar across subgroups? Should there be separate recommendations for subgroups? **Requirement for modeling:** Is there a lot of extrapolation and modeling required for these outcomes?
Values and preferences *Is there certainty or similarity?* The more certainty or similarity in values and preferences, the more likely a strong recommendation.	Yes ☒ No ☐	Benefits much higher valued than expected minor harms	A high value was placed on avoiding hospitalizations and mortality as well as improving quality of life A low value was placed on possible adverse events.	**Perspective taken:** Patient **Source of values:** Guideline panels assessment **Source of variability if any:** Not a lot of variability **Method for determining value; satisfactory for this recommendation:** Yes, given the expected small variability and difference between guideline panel and patients.
Resource implication *Are the resources worth the expected net benefit?* The lower the cost of an intervention compared to the alternative, and other costs related to the decision – that is, the fewer resources consumed – the more likely is a strong recommendation in favour of that intervention.	Yes ☒ No ☐	Resources required are worth the net benefit considering the benefit on mortality and hospitalizations	There are resources required to provide pulmonary rehabilitation to mortality and hospitalizations but these resources are worth the expected benefits and downstream treatment costs for cervical cancer are avoided. The treatment of adverse outcomes is also considered worthwhile.	**What are the cost per resource unit?** Although not evaluated here, a hospital bed per day is typically considered to be $800. Rehabilitation cost are approximately $3,000 to 5,000 per program per patient. **Feasibility:** Is this intervention generally available? **Opportunity cost:** Is this intervention and its effects worth withdrawing or not allocating resources from other interventions **Differences across settings:** Is there lots of variability in resource requirements across settings?
Overall strength of recommendation	Strong	The guideline panel recommends that patients with recent exacerbations of their COPD undergo pulmonary rehabilitation. (NOTE: this is a hypothetical recommendation developed for this article and not intended for clinical decision making.)		
Remarks and value- and preference statement	This recommendation places a high value on the benefits that can be expected (mortality reduction, reduction in hospitalizations and improvement in quality of life) and a relatively low value on the required resources. All patients should receive recommended usual care in addition to rehabilitation. (NOTE: this is a hypothetical recommendation developed for this article and not intended for clinical decision making.)			

The next slide shows a slightly more detailed table relating to the same effort of moving from evidence to recommendations. In the very right hand column now are explanations provided that guideline panels can use to make these judgments. There are sub-domains that influence the various decision domains that were already shown on the previous slides. Depending on the process that a guideline panel may use, one or the other format of the table may be appropriate for taking the panel through the decision-making process. The sub-domains just provide the individual decision or consideration criteria that panels should have in mind when they make this decision. At the end, the panel formulates a recommendation and provides information about what assumptions were made when making this recommendation.

Group composition

- Group composition might affect recommendation
- Common principle:
  include all affected by the recommendations
  (➔ multi-disciplinary groups incl. patients/carers) – Industry?
- Keep a manageable size

McMaster University
Inspiring Innovation and Discovery

The Process:
How to make it constructive?

- Group members are heterogeneous and might have different objectives
- Chair facilitates rather than leads the group
- Common understanding of goal, tasks and ground rules
- Similar level of required knowhow and skills
- Sufficient technical support

McMaster University
Inspiring Innovation and Discovery

Balanced participation and formal agreement

- Key task of chair

- Formal consensus processes
  - Delphi Method
  - Nominal group process
  - Voting

Group processes

Consensus development method	Mailed questionnaires	Private decisions elicited	Formal feedback of group choices	Face-to-face contact	Interaction structured	Aggregation method
Informal	No	No	No	Yes	No	Implicit
Delphi method	Yes	Yes	Yes	No	Yes	Explicit
NGT	No	Yes	Yes	Yes	Yes	Explicit
RAND version	Yes	Yes	Yes	Yes	Yes	Explicit
Consensus development conference	No	No	No	Yes	No	Implicit
Other methods						
Staticised group	No	Yes	No	No	–	Explicit
Social judgement analysis	No	Yes	Yes	Yes	No	Implicit
Structured discussion	No	No	No	Yes	Yes	Implicit

How to present controversies

- Lay out the controversies
- Describe the evidence
- Ask members to focus on the agreed upon evidence and the factors leading to a decision
- Ask whether there still is disagreement
- Vote
  - Make voting explicit and transparent (ways of doing this to come tomorrow)

Conclusions - Process

- Success depends on strong chair(s), training of group, good facilitation and technical support
  - Clinical and methods co-chairs
- Formal consensus developing methods might support agreement on recommendations
  - Voting represents forced consensus
- Guideline development will require sufficient resources.

McMaster University
Inspiring Innovation and Discovery

GRADE Grid

GRADE grid for recording panellists' views in development of guidelines (including examples of propositions from the Surviving Sepsis Campaign and number of panellists who voted for each option)

	GRADE score				
	1	2	0	2	1
Balance between desirable and undesirable consequences of intervention	Desirable clearly outweigh undesirable	Desirable probably outweigh undesirable	Trade-offs equally balanced or uncertain	Undesirable probably outweigh desirable	Undesirable clearly outweigh desirable
Recommendation	Strong: "definitely do it"	Weak: "probably do it"*	No specific recommendation	Weak: "probably don't do it"*	Strong: "definitely don't do it"
For each proposition below, please mark with an "X" the cell that best corresponds to your assessment of the available evidence, in terms of benefits versus disadvantages					
Use of (as opposed to no use of):					
Low dose steroids in patients with septic shock responsive to fluids and vasopressors	0	5	4	8	4
Low dose steroids in patients with septic shock poorly responsive to fluids and vasopressors	5	16	0	0	0
SDD in ventilated patient (local and systemic)	0	9	4	8	1
rhAPC in patients with septic shock and high risk of death	6	15	1	0	0

SDD=selective digestive decontamination, rhAPC= recombinant human activated protein C.

*Participants were provided with guidance on factors to be taken into account in formulating a recommendation (box 1) and the implications of strong versus weak recommendations (box 2).

GRADE Grid

GRADE grid for recording panellists' views in development of guidelines (including examples of propositions from the Surviving Sepsis Campaign and number of panellists who voted for each option)

	GRADE score				
	1	2	0	2	1
Balance between desirable and undesirable consequences of intervention	Desirable clearly outweigh undesirable	Desirable probably outweigh undesirable	Trade-offs equally balanced or uncertain	Undesirable probably outweigh desirable	Undesirable clearly outweigh desirable
Recommendation	Strong: "definitely do it"	Weak: "probably do it"*	No specific recommendation	Weak: "probably don't do it"	Strong: "definitely don't do it"
For each proposition below, please mark with an "X" the cell that best corresponds to your assessment of the available evidence, in terms of benefits versus disadvantages					
Use of (as opposed to no use of):					
Low dose steroids in patients with septic shock responsive to fluids and vasopressors	0	5	4	8	4
Low dose steroids in patients with septic shock poorly responsive to fluids and vasopressors	5	16	0	0	0
SDD in ventilated patient (local and systemic)	0	9	4	8	1
rhAPC in patients with septic shock and high risk of death	6	15	1	0	0

SDD=selective digestive decontamination, rhAPC= recombinant human activated protein C.
*Participants were provided with guidance on factors to be taken into account in formulating a recommendation (box 1) and the implications of strong versus weak recommendations (box 2).

An alternative method to formulating recommendations is shown here. The GRADE grid for voting of recommendations.

Formulate question

Select outcomes

Rate importance

Outcomes across studies

Create evidence profile with GRADEpro

Rate quality of evidence for each outcome

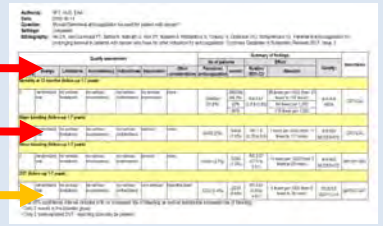
P
I
C
O

Outcome Critical

Outcome Critical

Outcome Important

Outcome Not important



Summary of findings & estimate of effect for each outcome

High
Moderate
Low
Very low

Randomization increases initial quality

Grade down

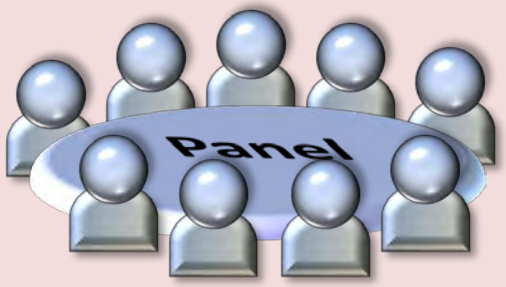
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

Grade up

1. Large effect
2. Dose response
3. Confounders

Systematic review

Guideline development



Grade
overall quality of evidence
across outcomes based on
lowest quality
of ***critical*** outcomes

Formulate recommendations:

- For or against (direction)
- Strong or conditional (strength)

By considering:

- ❑ Quality of evidence
- ❑ Balance benefits/harms
- ❑ Values and preferences

(Revise by considering:)

- ❑ Resource use (cost)

- "We recommend using…/should"
- "We suggest using…/might"
- "We recommend against using…/might not"
- "We suggest against using…/should not"

Agenda

McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

Agenda

McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**

Agenda

McMaster University
Inspiring Innovation and Discovery

09.00 h — 09.15 h **Welcome and introductions**
09.15 h — 10.30 h **Overview of the GRADE approach and process (large group)**
10.30 h — 10.45 h **Break**
10.45 h — 12.00 h **Assessing the quality of evidence (large group)**
12.00 h — 12.45 h **Break**
12.45 h — 14.30 h **Introduction to GRADEpro software, asking a question, specifying outcomes, grading quality of evidence (small group, hands-on)**
14.30 h — 15.00 h **Developing recommendations (large group)**
15.00 h — 15.15 h **Break**
15.15 h — 16.00 h **Developing recommendations (small group, hands-on)**
16.00 h — 17.00 h **Issues, challenges, questions, feedback**